

Análisis y estudio de la red de publicidad online de Yahoo

Trabajo Fin de Grado desarrollado por:

Beatriz Martínez Simón

Grado en Ingeniería de Sistemas Audiovisuales

Tutor: Juan Miguel Carrascosa Amigo

UC3M, Junio 2015



Universidad
Carlos III de Madrid

Índice

1.	Introducción	6
2.	Motivación	8
3.	Objetivos	10
4.	Publicidad en Internet	11
4.1.	Publicidad en Internet: Tipos.....	13
4.2.	Publicidad dirigida	14
4.3.	Beneficios de la publicidad dirigida	16
4.4.	Privacidad y seguridad.....	16
4.5.	Empresas que emplean publicidad dirigida	18
5.	Yahoo Advertising	20
5.1.	Recogida de información	20
5.2.	Categorías de intereses: Ad Interest Manager	21
5.3.	Exclusión de la publicidad personalizada.....	24
5.4.	Privacidad de los usuarios de Yahoo.....	25
5.4.1.	<i>Menores de 13 años</i>	<i>26</i>
5.4.2.	<i>Información sobre la salud</i>	<i>26</i>
5.4.3.	<i>Obtención de datos demográficos</i>	<i>26</i>
5.4.4.	<i>Compartición de información</i>	<i>27</i>
5.4.5.	<i>Inhabilitación de la publicidad</i>	<i>27</i>
5.4.6.	<i>Cookies</i>	<i>27</i>
6.	Planificación	29
7.	Desarrollo de la herramienta	30
7.1.	Tecnologías utilizadas.....	31
7.1.1.	<i>Bloque de automatización: Automatización de las búsquedas con iMacros.....</i>	<i>31</i>
7.1.1.1.	<i>¿Qué es iMacros?</i>	<i>32</i>
7.1.1.2.	<i>Utilización de iMacros en la herramienta</i>	<i>32</i>
7.1.2.	<i>Bloque de control y procesado: Shell Scripting</i>	<i>32</i>
7.1.2.1.	<i>¿Qué es un Shell?</i>	<i>32</i>
7.1.2.2.	<i>Shell Scripting</i>	<i>33</i>
7.1.2.3.	<i>Utilización de Shell Scripting en la herramienta.....</i>	<i>33</i>

7.1.3.	<i>Bloque de almacenamiento: Gestión de la información con MySQL</i>	33
7.1.3.1.	<i>Bases de datos en MySQL</i>	33
7.1.3.2.	<i>Utilización de Mysql en la herramienta</i>	34
7.2.	<i>Bloque de configuración</i>	34
7.3.	<i>Bloque de almacenamiento: Acceso 1</i>	35
7.4.	<i>Bloque de control y procesado: Acceso 1</i>	39
7.5.	<i>Bloque de automatización</i>	40
7.6.	<i>Bloque de control y procesado: Acceso 2</i>	45
7.7.	<i>Bloque de almacenamiento: Acceso 2</i>	46
8.	<i>Batería de experimentos</i>	47
8.1.	<i>Experimentos realizados con Correo Yahoo</i>	48
8.1.1.	<i>Definición de los experimentos</i>	48
8.1.2.	<i>Ejecución de los experimentos</i>	48
8.1.3.	<i>Observaciones</i>	50
8.2.	<i>Experimentos realizados con Flickr</i>	50
8.2.1.	<i>Definición de los experimentos</i>	50
8.2.2.	<i>Ejecución de los experimentos</i>	50
8.2.3.	<i>Observaciones</i>	51
8.3.	<i>Experimentos realizados con Yahoo Screen</i>	51
8.3.1.	<i>Definición de los experimentos</i>	51
8.3.2.	<i>Ejecución de los experimentos</i>	52
8.3.3.	<i>Análisis de los resultados</i>	52
8.3.4.	<i>Observaciones</i>	58
8.4.	<i>Experimentos realizados con Buscador Yahoo</i>	59
8.4.1.	<i>Definición de los experimentos</i>	59
8.4.2.	<i>Ejecución de los experimentos</i>	59
8.4.3.	<i>Análisis de los resultados</i>	60
8.4.4.	<i>Observaciones</i>	67
8.5.	<i>Comparativa Buscador y Yahoo Screen</i>	68
9.	<i>Conclusión</i>	70
ANEXO I	72
ANEXO II	73
ANEXO III	74
ANEXO IV	76

ANEXO V.....	77
ANEXO VI.....	78
Referencias.....	79

1. Introducción

La sociedad de la información es aquella en la que las tecnologías que facilitan la creación, distribución y manipulación de la información, juegan un papel esencial en las actividades sociales, culturales y económicas. [1]

No nos equivocamos al conceder que formamos parte de esta sociedad, ya que se podría decir que prácticamente todo funciona a través de Internet, pues está totalmente integrado en nuestro quehacer diario y ya hay una generación para la que juega un papel esencial en su vida cotidiana. Hay que tener en cuenta, además, que gracias a las bajas barreras de entrada del mundo digital, los medios se multiplicaron y fragmentaron, creando así el extenso ecosistema digital del que disponemos hoy.

No es de extrañar, por tanto, que incluso la publicidad, cuyo propósito en la actualidad no sólo es vender productos y promocionar servicios [2], sino que también se emplea para difundir información, se haya hecho con un hueco importante dentro de la red, reportando numerosos beneficios económicos en todo el mundo [3]:

- En 2011 generó 86,4 billones de dólares. En 2013 se incrementaron los ingresos en un 36%, alcanzando los 117,6 billones de dólares.
- Para 2016 se estima que el presupuesto en EEUU para publicidad doblará al de 2011 (alcanzando los 67,4 billones de dólares anuales).
- En la red social Youtube, el anuncio más compartido en la red en 2013 alcanzó a 4,24 millones de usuarios. El más visto se reprodujo 114 millones de veces.
- Un anuncio en la página principal de Youtube cuesta 400.000\$.
- El 87% de los ingresos en Twitter provienen de la publicidad.
- La media de CTR (*Click-Through-Rate*) en anuncios de Facebook aumentó un 275% de 2012 a 2014.
- Un 95% de los ingresos totales de Google provienen de la publicidad.
- En otoño de 2012, tenía más de 1,2 millones de anuncios de empresas en su *Search Network*.

A tenor de estos datos, se puede establecer como una realidad el hecho de que la publicidad digital tiene un papel importante en la sociedad actual, reportando millones de dólares en todo el mundo, ya no solo a los anunciantes, sino también a los sitios web que dan soporte a este tipo de publicidad, y que actúan a su vez como medio de transmisión, y a las agencias encargadas de proporcionar los servicios publicitarios.

Es por todo esto que resulta especialmente interesante conocer más sobre la publicidad digital: cómo funciona, qué estrategias sigue, qué supone para los agentes que interactúan con ella, qué repercusiones tiene en la sociedad, etc.

Así pues, a lo largo de esta memoria, nos adentraremos en el mundo de la publicidad digital, prestando especial atención a la publicidad basada en intereses que, como veremos, se trata de un tipo de publicidad dirigida a usuarios particulares, y trataremos contestar a las preguntas que surgen sobre su funcionamiento y sus resultados desde el punto de vista del usuario. Para esto, plantearemos el estudio centrado en el sitio web de Yahoo, en el que, a través de los diferentes servicios que proporciona, trataremos de determinar cómo establece qué le interesa a cada uno de sus usuarios y así, ofrecerles publicidad personalizada, de modo que la experiencia publicitaria resulte un tanto más beneficiosa tanto para los usuarios como, también, a los anunciantes.

2. Motivación

Ya que hemos hablado de la *sociedad de la información*, podemos decir que, en todas las áreas, la creación, el almacenamiento, el control, el transporte o la difusión de la información tienen ahora una valiosa importancia estratégica y es uno de los sectores económicos fundamentales en los países desarrollados, ya que contribuye decisivamente a mejorar la productividad y permite acceder a mercados más amplios. Al ser el instrumento fundamental de comunicación de ámbito mundial, se ha convertido en un recurso público del que dependen empresas, administraciones y personas en general [4].

Más allá del concepto de *sociedad de la información*, es indiscutible que, en pleno siglo XXI y en medio de una importante revolución tecnológica, podemos decir con libertad que vivimos en una sociedad *pro-consumo* [5]. Ante esto, sabemos que la publicidad juega un papel realmente importante en nuestra vida cotidiana y, si a esto le sumamos el concepto anterior, podemos concluir en que la publicidad en Internet es un artificio muy valioso para llegar a las personas.

Es por esto que las grandes empresas de Internet han desarrollado un método publicitario denominado Online Advertising o, como lo llamaremos aquí, Publicidad en Internet, que es una rama de la publicidad digital que se centra en llegar a los posibles consumidores de sus productos a través de la red.

Una forma específica de este tipo de publicidad, y que se ha puesto muy de moda en las redes sociales, es la publicidad basada en intereses, cuyo objetivo es ofrecer a los usuarios de Internet, anuncios que se ajustan a sus gustos y/o necesidades.

La publicidad basada en intereses se desarrolla con el fin de crear un modelo publicitario que sea beneficioso tanto para el anunciante como para la persona que vea ese anuncio, es decir, el objetivo principal es, del lado de los anunciantes, segmentar de forma más precisa a sus potenciales clientes y así abaratar costes, y, del lado de los usuarios, no ser bombardeados con anuncios de productos que no tienen ningún interés para ellos, sino visualizar anuncios publicitarios que puedan resultarles interesantes de algún modo.

Para llevar a cabo un buen ejercicio de este tipo de publicidad, se ha de recopilar información de los usuarios mediante diferentes vías, ya sea a través de cookies, información recopilada a través de sitios web externos sobre la actividad de los usuarios o a través de información y *keywords* de búsqueda de los mismos. Esta información que se recoge de los usuarios es tratada con el fin de determinar qué tipos de intereses tiene un usuario concreto para así poder mostrarle anuncios relacionados con algo que le pueda interesar, de esta forma es más probable que también salga beneficiado el anunciante, pues si el anuncio se muestra a un cliente potencial, es más

probable que se realice una venta que si el anuncio se mostrase aleatoriamente a cualquier persona. Por supuesto, ante este tipo de publicidad habrá que tener en cuenta las políticas de privacidad que se aplican y las leyes vigentes sobre la protección de datos de los usuarios.

De esta forma, estudiaremos el sistema que desarrolla un sitio web concreto (en este caso será Yahoo) tanto a la hora de recopilar información de sus usuarios como a la hora de categorizar los intereses de estos antes de desarrollar su sistema publicitario. Es en este último punto, cuando se establecen los intereses a través de la información recopilada, donde surgen más dudas, pues no tenemos información de cómo funciona realmente este sistema, así que necesitaremos desarrollar una herramienta automatizada y fácilmente configurable para poder interactuar con los servicios de Yahoo de forma automática y así, mediante distintos experimentos, poder aproximar una idea de lo que está sucediendo en realidad en un sistema de publicidad basada en intereses.

3. Objetivos

Como hemos venido diciendo, a lo largo de este trabajo realizaremos un estudio de la publicidad basada en intereses, utilizando como "sujeto de pruebas" el sitio web Yahoo, que dispone de un servicio de publicidad de este tipo.

Para ello, llevaremos a cabo el desarrollo de una herramienta que interactúe con diferentes servicios proporcionados por Yahoo (Buscador, Correo, Flickr y Screen), obteniendo así diferentes datos sobre la forma de asignación de preferencias. LA herramienta será completamente configurable para poder generar experimentos adaptados a nuestras necesidades. A partir de los datos obtenidos, que posteriormente analizaremos, intentaremos determinar el funcionamiento y la lógica de la asignación de etiquetas sobre los gustos de los usuarios de Yahoo.

A continuación listamos los objetivos planteados para este proyecto:

- Analizar el ecosistema de la publicidad digital y analizar los distintos tipos de estrategias usadas para desarrollar este tipo de publicidad.
- Estudiar y analizar los servicios que ofrece Yahoo y cómo influyen en su modelo de publicidad personalizada, así como profundizar este estudio para los siguientes servicios: Yahoo Screen, Flickr, Correo de Yahoo y Buscador Yahoo.
- Diseñar una herramienta que nos permita realizar el estudio anterior de una forma flexible y automatizada.
- Plantear una batería de experimentos, que configuraremos y lanzaremos con la herramienta, y que nos permitan obtener datos útiles sobre las hipótesis que realiza Yahoo para personalizar la publicidad de sus usuarios.
- Procesar y realizar un análisis de los datos que proporciona la herramienta.
- Evaluar los resultados obtenidos a fin de intentar comprender el mecanismo que utiliza Yahoo en sus diferentes servicios para determinar los intereses de los usuarios y así proporcionarles publicidad personalizada.

4. Publicidad en Internet

La publicidad online es la publicidad que se realiza en los medios interactivos: Internet, televisión interactiva y aparatos móviles. Con esta publicidad, los anunciantes pretenden conocer información de los usuarios, así como sus intereses y sus necesidades, con el fin de alcanzar a clientes potenciales para vender sus productos [6].

Antes de llegar hasta este concepto, el ecosistema de la publicidad estaba formado únicamente por tres actores: el anunciante, la agencia y el medio. Después, este modelo evolucionó hasta alcanzar la estructura de la que dispone en la actualidad, en la que participan múltiples actores, muchos de ellos con un enfoque tecnológico, para así adaptar la industria a los tiempos [7].

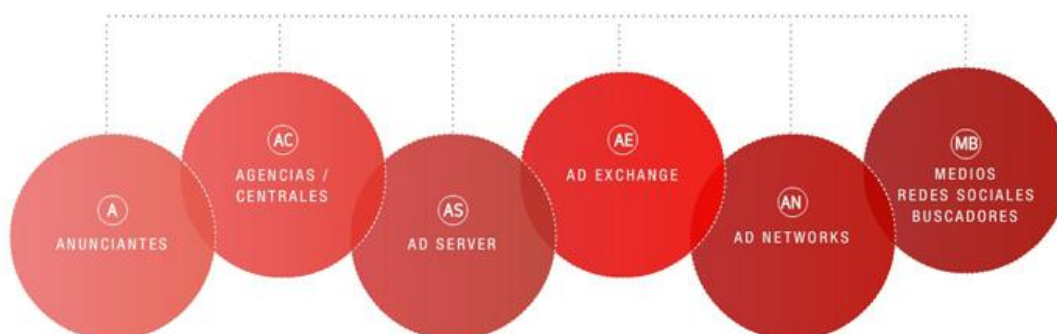


Figura 1. Distintos actores que conforman el actual ecosistema de la publicidad digital [7].

Tras ver este ecosistema, existen algunos conceptos básicos que necesitamos conocer para hablar de la publicidad online.

El primero es el medio o *publisher*, que es la página web donde se coloca la publicidad en cuestión, pueden ser redes sociales, blogs, foros, páginas de empresas, buscadores, periódicos digitales, etc. [8].

El siguiente concepto es el de anunciantes o *advertisers* que, como su propio nombre indica, son las empresas o particulares que desean colocar los anuncios en los distintos soportes a fin de captar clientes [8].

Para llevar a cabo la interacción de los anunciantes con el medio, necesitamos a los actores que se muestran en la Figura 1 como *Ad Server*, *Ad Exchange* y *Ad Networks*, que definimos a continuación.

El *Ad Server* [9] es una herramienta digital encargada de emitir los anuncios sobre los distintos soportes. El *Ad Exchange* [10] es la plataforma digital que se encarga de la compra y la venta de los espacios publicitarios. Por último, el *Ad Networks* [11] es lo que definiríamos como agencia, pues se encarga de ofrecer el servicio publicitario a los

anunciantes y de colocarlo en los soportes con las herramientas que ya hemos definido para ello.

Una vez tenemos la publicidad, entran en juego los consumidores o *consumers*, que son los visitantes de las páginas web y clientes potenciales de los anuncios que hay distribuidos en la página que visitan. No obstante, hablamos de clientes potenciales porque sólo un porcentaje pequeño de los que vean un anuncio harán clic en él. Para explicar esto, definiremos los conceptos de CTR y CR [8].

El CTR (*Click Through Rate*) es el número de clics que se producen cada 1.000 impresiones, es decir, cada 1.000 anuncios mostrados, expresado en tanto por ciento. Si un usuario hace clic en un anuncio, accede a la página de venta (*landing page*), pero no todo el que entre allí será un cliente. El CR (*Conversion Rate*) es el número de ventas por cada 1.000 visitas, expresado en tanto por ciento.

Estos conceptos de medida dentro de la publicidad en Internet son útiles para calcular los costes que conllevan a los anunciantes el hecho de generar publicidad online. Hay varios tipos de medida de estos costes, que se resumen en la Tabla 1.

Con la cantidad de sitios web que tenemos a nuestro alcance y toda la publicidad que se nos presenta en ellos, deducimos que puede resultar realmente difícil, para un anunciante, conseguir una venta. Es por este motivo por lo que entran en juego las

Tipo de medida	Definición
CPA (Cost Per Action)	Coste pagado por acción (clic, venta o registro).
CPC (Cost Per Click)	Coste pagado por clic.
CPL (Cost Per Lead)	Coste total de la campaña de publicidad en línea, dividido entre el número de clientes potenciales generados.
CPM (Cost Per thousand impressions)	Coste pagado por cada mil impresiones (visualizaciones) de un anuncio.
CPS (Cost Per Sale)	Coste pagado por venta generada.

Tabla 1. Glosario de tipos de medidas de costes en la publicidad digital [12].

agencias de publicidad en Internet, encargadas de colocar la publicidad de forma estratégica para que resulte beneficiosa, no sólo para el anunciante, sino también para el usuario y potencial cliente. Es aquí donde toma parte el concepto de publicidad dirigida que, como veremos más adelante, se basa en mostrarse a usuarios que realmente puedan estar interesados en ella. Pero antes de centrarnos en ello, veremos a grandes rasgos los distintos tipos de publicidad en Internet.

4.1. Publicidad en Internet: Tipos

La publicidad en Internet (también llamado Publicidad Online u Online Advertising) es una forma de dirigirse a los potenciales clientes a través de Internet. Dentro de este tipo de publicidad se consideran varios tipos: email marketing, marketing en buscadores (SEM), marketing en redes sociales, posicionamiento web (SEO), Display Ads (publicidad gráfica como los banners) y, últimamente, la publicidad para teléfonos móviles.

Como podemos ver, hay muchos tipos de publicidad en Internet, pero el caso que estudiamos se centra en el bloque de publicidad estratégica dirigida y, más concretamente, en el Display Advertising, comprendido dentro del Search Engine Marketing (SEM). Es por esto que vamos a desarrollar más en profundidad estos conceptos.

EL SEM es una técnica de marketing diseñada para aumentar la visibilidad de un sitio web en los resultados de los motores de búsqueda (SERP). Se vale de un trabajo de optimización gratuito (SEO) y de uno de pago (Display Ads, Sponsored Search).

El Display Advertising transmite un mensaje publicitario utilizando texto, logos, animaciones, vídeos o fotos. Los Display Ads normalmente apuntan a usuarios de características específicas, buscando así la manera de aumentar el efecto de los anuncios. Existen tres técnicas mejoradas, diseñadas para aumentar el retorno de la inversión (ROI) [13]:

- a) Retargeting: las cookies pueden realizar un seguimiento de si un usuario deja una página sin comprar nada, por lo que el anunciante puede después hacer uso del Retargeting al usuario con los anuncios de la página que el usuario visitó sin finalizar la operación.
- b) Segmentación por Comportamiento (Behavioral Targeting): como los anunciantes recopilan datos a través de múltiples sitios web externos sobre la actividad en línea de un usuario, se puede crear una imagen detallada de los intereses del usuario para ofrecer la publicidad más dirigida. Esto es la llamada Segmentación por Comportamiento (Behavioral Targeting).
- c) Publicidad Semántica y Contextual: publicidad que trabaja con información y *keywords* de búsqueda del usuario, por lo que normalmente ofrece contenido relacionado con lo visto previamente por el usuario. Por ejemplo, si a un usuario le gusta el fútbol y ha visto páginas de esto, las páginas que tengan espacio para publicidad contextual mostrarán en ese espacio información de páginas de fútbol, entradas para los partidos, etc.

Técnica	Definición
Retargeting	Técnica que utiliza la información de las cookies para detectar qué páginas abandona el usuario sin comprar y poder mostrarle publicidad de ello más tarde.
Segmentación por Comportamiento (Behavioral Targeting)	Técnica que recopila información de los usuarios a través de sitios web externos para determinar sus intereses.
Publicidad Semántica y Contextual	Técnica que ofrece contenido relacionado con lo visto previamente por el usuario a partir de la información y <i>keywords</i> de búsqueda del usuario.

Tabla 2. Glosario de las técnicas diseñadas para aumentar el retorno de la inversión en la publicidad digital.

El caso que nos ocupa funciona con la segmentación por comportamiento (Behavioral Targeting), como veremos más específicamente en el apartado 5.

4.2. Publicidad dirigida

La publicidad basada en intereses, también conocida como publicidad personalizada o IBA (Interest Based Advertising), consiste en mostrar anuncios publicitarios que se esperan sean relevantes para los intereses de cada usuario en particular. De esta forma, la publicidad basada en intereses permite a los anunciantes llegar a los usuarios en función de sus intereses particulares, suponiéndose que, de este modo, las personas que vean un anuncio pueden estar interesadas en él de algún modo [14].

Vamos a explicar esto un poco mejor, de forma gráfica y apoyándonos en el ejemplo representado en Figura 2.

Esta figura representa un ejemplo de publicidad dirigida en dispositivo móvil. Como podemos ver, se muestran varios días de la semana, en los que el usuario interactúa con los anuncios que se le muestran en su dispositivo y, en función de lo que éste haga, el sistema podrá determinar una categoría de interés ajustada al usuario.

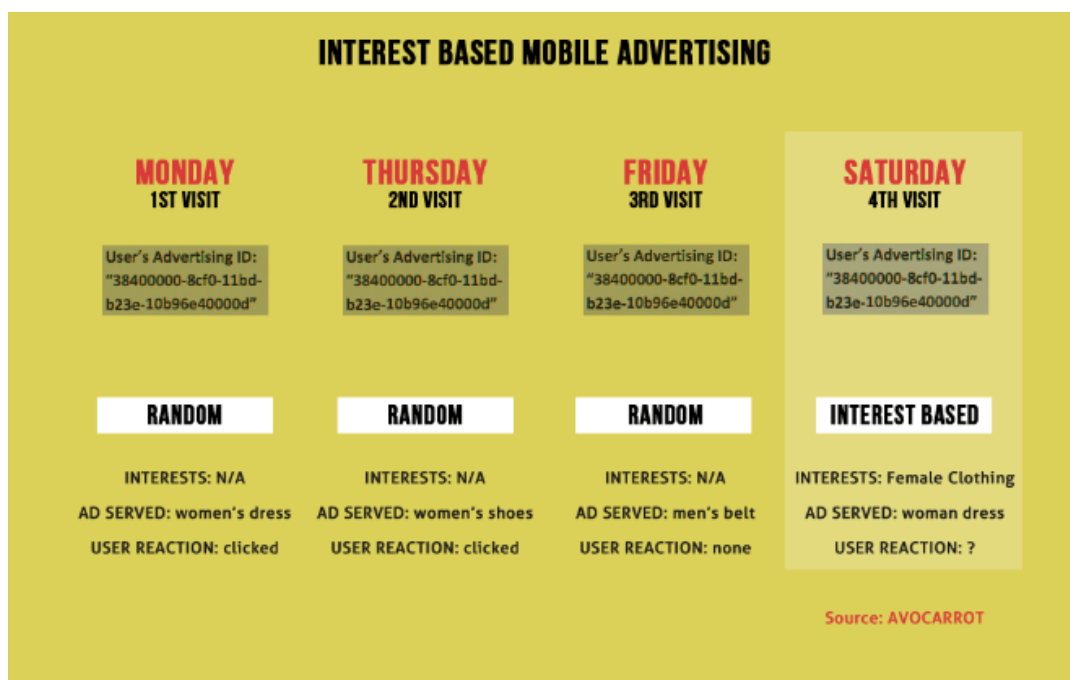


Figura 2. Ejemplo de publicidad dirigida [15]. Se muestra el proceso que se seguiría al mostrar publicidad online, en un dispositivo móvil, a un usuario al que se le asigna como interés la ropa de mujer.

	Intereses almacenados	Anuncio mostrado	Naturaleza del anuncio	Reacción del usuario
Lunes	--	Vestido mujer	Aleatoria	Clic
Jueves	--	Zapatos mujer	Aleatoria	Clic
Viernes	--	Cinturón hombre	Aleatoria	Ignorado
Sábado	Ropa mujer	Vestido mujer	Basado en intereses	?

Tabla 3. Ejemplo de publicidad dirigida representado en la figura 1.

Tal y como se muestra en la Tabla 3, el sistema mostraría anuncios aleatorios al usuario y, a partir de sus reacciones ante ellos (si los ignora o no), sería capaz de determinar qué tipo de anuncios le interesan, como en este caso, la ropa de mujer.

En general, se envía publicidad basada en intereses mediante el uso de una cookie anónima u otro identificador asignado a su dispositivo. Dichas cookies u otros identificadores se ubican dentro de unas categorías de intereses generales, basándose en cierta información, como por ejemplo:

- Las páginas web que el usuario visita y los enlaces en los que hace clic.
- Las aplicaciones móviles que usa.
- Algunos correos electrónicos.

- Los datos demográficos.
- Datos de compra en línea y fuera de línea.

Normalmente, la página web que proporciona esta publicidad basada en intereses confía en terceros para que les ayuden a enviar a los usuarios dicha publicidad y permiten que dichos terceros puedan recopilar información sobre sus actividades en línea, a través de las cookies y de otras tecnologías que, no solamente sirve para mostrar anuncios de interés para el usuario, si no que tiene otras aplicaciones como el no mostrar repetidamente el mismo anuncio.

4.3. Beneficios de la publicidad dirigida

Con este sistema de publicidad, los anuncios en línea que el usuario recibe se adaptan mejor a sus intereses. Esto significa que es más probable que vea los anuncios de los productos que le interesan.

Una funcionalidad que suele incorporarse es que el usuario no vea los mismos anuncios una y otra vez, dado que se limita la cantidad de veces que se ve un anuncio.

Al mismo tiempo, la publicidad basada en intereses debería ayudar a obtener ingresos de los sitios web donde se utiliza este sistema de forma más eficiente, a aumentar el valor para los anunciantes y a proporcionar una mejor experiencia a los usuarios.

4.4. Privacidad y seguridad

No obstante, es incuestionable que este sistema publicitario se adentra en un terreno peligroso como es el de la privacidad de los usuarios.

A la hora de hablar de publicidad personalizada, para la que es necesario recabar información personal del usuario a través de sus acciones en distintos sitios de Internet, debe tenerse en cuenta que la privacidad es un concepto vago, que se caracteriza por tener la capacidad de variar geográfica, temporal, cultural y generacionalmente.

Partiendo de esta característica, podemos decir que el concepto de privacidad será un problema mayor o menor dependiendo de en qué entorno nos encontremos.

Por ejemplo, las cookies son archivos que usan las webs para almacenar información sobre los usuarios. Pero las cookies de terceros tienen la

particularidad de que no están ligadas a la página que el internauta visita, por lo que acumulan muchos datos privados relacionándolos entre sí sin su consentimiento. Esto produce un gran malestar entre los usuarios, ya que consideran que con estas prácticas su privacidad corre peligro.

En la Unión Europea se dispone de una Directiva de protección de datos sobre la que se han ajustado las legislaciones de los quince países miembros, a fin de que todos los ciudadanos cuenten con una protección equivalente. Además, se dispone de otra Directiva que se ocupa específicamente de la protección de la intimidad en el sector de las telecomunicaciones [16] .

No obstante, y aunque las leyes de protección de datos de cada país se ajusten a la Directiva de la UE, hay diferencias entre ellas. Vemos un ejemplo de esto en la Figura 3 [17], donde se muestra una comparativa entre cinco países miembros de sus leyes de protección de datos y el código penal aplicado en cada caso.

País	Ley de Protección de Datos		Código penal
España	Arts. 44-45 Infracciones leves: 601,01 a 60.101,21 € Infracciones graves: 60.101,21 a 300.506,05 € Infracciones muy graves: 300.506,05 a 601.012,10 €		1 a 4 años prisión + multa de 12 a 24 meses - tipo básico (Art. 197.2 CP) 2 a 5 años prisión - tipo agravado 3 a 5 años prisión - tipo agravado encargado o responsable 4 a 7 años prisión- tipo hiperagravado
Alemania	Párrafo 43 Multa de 25.564 a 255.645 €	Párrafo 44 Hasta 2 años de cárcel o multa por las conductas del párrafo 43 realizadas con ánimo de lucro o con intención de perjudicar a un tercero	Párrafo 202a Espionaje de datos: pena de prisión o multa no superior a 3 años
Francia	Art. 45 a 47 Primera infracción: multa no superior a 150.000 € Infracción reiterada: multa no superior a 300.000 €		Arts. 226-16 a 226-24 Pena de prisión de 5 años + multa de 300.000 €
Italia	Arts. 161-165 De 500 a 90.000 €	Arts. 167 a 172 Penas de 3 meses a 3 años	
Suecia	Multa de 6 meses a 2 años (sistema días multa)	Art. 49 Pena de 6 meses a 2 años	

Figura 3. Comparativa de la Ley de Protección de Datos en distintos países de la UE.

Es importante que se traten debidamente estos problemas y sus posibles soluciones, ya que el respeto de la privacidad y la protección de los datos personales son derechos fundamentales de la UE. Por ese motivo, es importante contemplar la necesidad de que los usuarios puedan optar por no recibir publicidad basada en intereses. Como veremos más adelante, Yahoo contempla esa opción y, en ese caso, los sistemas automatizados de Yahoo no explorarán o analizarán el contenido de las comunicaciones enviadas y recibidas desde la cuenta de dicho usuario (incluyendo Correo Yahoo y contenido Messenger) para tales fines.

Profundizaremos en el tratamiento de la privacidad y la seguridad de los usuarios por Yahoo en el apartado 5.4.

4.5. Empresas que emplean publicidad dirigida

Son muchas, y cada día más, las empresas que utilizan este tipo de publicidad en Internet. Un ejemplo de ello es Twitter, que por defecto muestra publicidad basada en el historial de los usuarios, de forma que en el *timeline* aparecen anuncios sobre las páginas que se hayan visitado. Claro está que, en un estudio más amplio de la actividad del usuario, también se basa en búsquedas realizadas o anuncios sobre los que se ha hecho clic anteriormente. Por ejemplo, como veíamos en el apartado 4.2, si el sistema detecta que el usuario ha hecho clic en varios anuncios de ropa de mujer, determinará que hay un interés en ese tema y le ofrecerá publicidad relacionada. Hay muchas formas de recabar información para personalizar la publicidad, que veremos más adelante en el apartado de *Yahoo Advertising*. No obstante, este tipo de publicidad es contextual, y este trabajo pretende centrarse en la segmentación por comportamiento.

También Facebook, a partir del 1 de Enero de 2015, ha introducido mejoras en sus anuncios en función de las aplicaciones y los sitios que sus usuarios utilizan fuera de la propia red social (publicidad basada en el comportamiento en Internet) y ha empezado a ampliar el control que cada usuario tiene sobre este tipo de publicidad [18].

Tanto Yahoo como Google, dos de los grandes colosos de Internet, disponen de un servicio de publicidad basada en intereses. Sin embargo, cada uno utiliza sus propios sistemas y herramientas para determinar qué anuncios muestran a los usuarios, aunque ambos poseen un concepto común, y es poder llegar a los usuarios en función de los intereses y los datos demográficos deducidos, además, claro está, de amortizar con ello sus servicios.

En general, los administradores de estos anuncios (como Google AdSense [19] o Yahoo Adbrite, entre otros) a menudo utilizan cookies (identificadores de usuario) y así determinan qué anuncios mostrar a cada consumidor en particular [13].

En el caso de Google podemos añadir que pretende revolucionar la publicidad dirigida con un sistema propio de rastreo de usuarios. Como hemos visto anteriormente, las cookies de terceros suponen un problema ante la privacidad de los usuarios, por ello, Google está trabajando en un nuevo modelo de rastreo. El sistema del que hablamos se llama AdID, y pretende cambiar la forma en que se obtienen los datos referidos a la navegación de los internautas. Este identificador anónimo pretende evitar que los anunciantes recopilen informaciones personales de los usuarios que éstos no quieren revelar [20].

El AdID permite que el usuario limite el seguimiento de su navegación a las empresas, pudiendo renovar cada año incluso la lista de anunciantes que aprueban o desaprueban, modificando esta opción en el navegador Chrome [20].

De Yahoo, que es nuestro sitio protagonista, podemos decir que cuenta con un sistema de predicción potente, que funciona mejor cuanto mayor es la actividad que realiza el usuario sobre los servicios de Yahoo.

Asimismo, Yahoo muestra las categorías de intereses estándar y proporciona al usuario la oportunidad de elegir si desea o no recibir anuncios personalizados de cada una de las categorías. De igual manera que ofrece la opción de descartar por completo la publicidad personalizada (dedicaremos el apartado 5.3 a hablar de esta opción de exclusión, también conocida como "Opt-Out"), en cuyo caso se mostrarían al usuario anuncios aleatorios de cualquier tema. Para la gestión de estas opciones, Yahoo dispone de una herramienta específica: el Ad Interest Manager.

En este trabajo haremos un estudio exhaustivo de cómo funciona la publicidad dirigida de Yahoo que, como hemos adelantado, se basa en la técnica del Behavioral Targeting que veremos, de forma detallada y especificada al caso de Yahoo, en el siguiente apartado.

5. Yahoo Advertising

En este proyecto nos centramos en el funcionamiento específico de la publicidad basada en intereses que utiliza Yahoo.

Para que los anuncios sean más relevantes y útiles para los usuarios, en Yahoo hacen hipótesis acerca de los intereses de sus usuarios, tomando como base su actividad en los sitios y servicios de Yahoo. A algunos de los anuncios que se muestran reflejan estos intereses, aunque no todos los anuncios publicitarios funcionan bajo este sistema pues, como se comentó en el apartado 4.2, el sistema puede necesitar observar el comportamiento del usuario frente a anuncios de temas aleatorios para poder determinar qué le interesa realmente.

Del mismo modo, cada usuario puede optar por no recibir publicidad basada en los intereses y el contenido recopilado a través del contenido de la exploración de las comunicaciones con fines publicitarios, el intercambio de información con los asociados, y análisis de recopilación de datos a través de los sitios asociados en conjunto con las herramientas de esta página.

La mayor parte de la información de detallamos a lo largo de este apartado está obtenida directamente de la información que proporciona Yahoo sobre su Ad Interest Manager [21].

5.1. Recogida de información

Para poder identificar los intereses de los usuarios, Yahoo utiliza la información sobre el uso que el cliente hace de sus servicios, que consigue por distintas vías [22].

- Cuando un usuario se registra para obtener una cuenta, Yahoo recoge información personal de dicho usuario.
- También se recoge información cuando se visitan sitios dentro de Yahoo o sitios de socios de Yahoo.
- A través la participación en promociones.
- Al realizar transacciones con Yahoo o alguno de sus socios. En este caso se recogen datos sobre la utilización de los servicios y productos financieros ofrecidos.
- Yahoo recibe y registra información proporciona el navegador: dirección IP, cookies de Yahoo y página solicitada.
- En las cookies puede almacenarse mucha información acerca del usuario: Las páginas que el usuario ha visitado recientemente en Yahoo, anuncios que ha visto y sobre los que ha hecho clic, los clic que desde otras páginas

web le han llevado hasta Yahoo, las búsquedas que se han realizado en Yahoo o el contenido de las comunicaciones escaneadas de las que el usuario ha dado el consentimiento.

Los datos obtenidos de cualquiera de estas maneras pueden combinarse con datos que socios de Yahoo, proveedores u otras empresas les proporcionan.

Yahoo usa dicha información para personalizar la publicidad y el contenido de acuerdo al usuario y mejorar sus productos, así como realizar estudios e investigaciones.

Centrándonos en la publicidad, esta información recabada entra en determinados sistemas algorítmicos que ayudan a hacer una conjetura acerca de los intereses del usuario en cuestión.

5.2. Categorías de intereses: Ad Interest Manager

Para que los anuncios sean más relevantes y útiles para cada usuario, Yahoo puede hacer conjeturas acerca de sus intereses en función de su actividad en nuestros sitios y servicios. Al utilizar Yahoo Mail, los sistemas automatizados de exploración y análisis detectan, entre otras cosas, ciertas palabras y frases (llamadas "palabras clave") dentro de sus comunicaciones y también del contenido enviado y recibido desde su cuenta. Además de usar las palabras clave para que a los usuarios se les muestre el contenido y los anuncios contextualmente relevantes, estas palabras clave también pueden contribuir a la comprensión de las cosas que le interesan a cada usuario. Estas categorías de interés se muestran en el *Ad Interest Manager* [23].

La categorización de intereses de Yahoo está impulsada por varios factores: las búsquedas, la propia actividad y la interacción con la publicidad. Como la categorización de intereses está la realiza un ordenador, el sistema trata de recoger las actividades colectivas de los usuarios y de equipararlas a un tema en particular.

A veces, el sistema recibe muy poca información para determinar con precisión unos intereses concretos o la información que se ofrece puede llevar a pensar que el usuario tiene múltiples intereses, basados en significados ambiguos a través de la información proporcionada. Cuantas más actividades se realizan a través de Yahoo, mejor es la capacidad de predicción de los sistemas de Yahoo para adivinar los intereses del usuario y aumentar la conveniencia de los contenidos y la publicidad que visualiza.

Para determinar qué categorías puede ver y editar el usuario, Yahoo muestra las categorías de intereses estándar de entre las que el usuario puede elegir para recibir anuncios personalizados. Las categorías de intereses se basan en un sistema de "scoring" (puntuación o tanteo en castellano) que categoriza muchas de las páginas vistas, clics en páginas y anuncios y las búsquedas llevadas a cabo. Cuando un usuario alcanza un cierto umbral, este es elegido para recibir los anuncios que están asociados con esas categorías de interés.

Además, los anunciantes solicitan a Yahoo crear categorías de interés personalizadas específicamente para ellos. Por ejemplo, es posible que un anunciante desee mostrar anuncios sólo a las personas que han visitado una página específica en su sitio o han buscado una palabra clave o palabras específicas. Estas categorías personalizadas no aparecen en el Ad Interest Manager de Yahoo.

El Ad Interest Manager (Figura 4) es una herramienta que permite a los usuarios de Yahoo indicar y almacenar opciones publicitarias para llevar a cabo una correcta personalización de la publicidad siempre que Yahoo ofrece anuncios. Es por esto, que se les permite a los usuarios eliminar categorías de forma manual. Una vez que se elimina una categoría de intereses, el usuario ya no recibirá publicidad personalizada para esa categoría. Existe un mecanismo de desactivación de este tipo de publicidad, la opción "Opt-Out", que se estudiará en el apartado siguiente.

Además, en el sistema de categorización de Yahoo existen medidas relativas de la actividad de cada usuario. Por ejemplo, si un usuario visita Yahoo Deportes con menos frecuencia que el visitante medio, se registrará como un visitante de frecuencia "baja". Si, por el contrario, lo visita con más frecuencia que el visitante medio, se le mostrará como un visitante de frecuencia "alta". Existen tres tipos de frecuencia: "baja", "media" y "alta".

Sería realmente interesante poder acceder a esta clasificación de la frecuencia que utiliza Yahoo, no obstante, no se muestra de forma pública, forma parte de su sistema y no podemos acceder a ella a través de los experimentos, por lo que tendremos que realizar nuestros estudios sin esta información.

¿Eres nuevo aquí? [Regístrate](#) | [Iniciar sesión](#) | [Help](#) | [Download the new Yahoo Mail app](#) | [Correo](#) | [Yahoo](#)

YAHOO! [Buscar en la Web](#)

[Página principal](#) [Productos](#) [Temas](#) [Preferencias](#) [Ayuda](#)

Ad Interest Manager [Enviar página](#) [Imprimir página](#)

Qué cubre Ad Interest Manager

Yahoo! aplicará tu cancelación a ciertos otros productos que ofrecemos, incluyendo el análisis de contenido de comunicaciones con propósitos de publicidad, para recibir contenido basado en intereses y para recibir información de sitios afiliados sobre nuestros productos analíticos.

Es tu Web!

Yahoo! es un sitio web patrocinado por anunciantes. Muchos de los productos y servicios que ofrecemos son gratuitos para el usuario, a cambio de poder mostrar publicidad en los mismos. Otros sitios web también se asocian con Yahoo! para mostrar publicidad que les permita patrocinar sus ofertas. Para más información, pincha aquí.

Ad Interest Manager

Para hacer nuestros anuncios más relevantes y útiles para ti, hacemos estimaciones razonadas sobre tus intereses basándonos en tu actividad en los sitios y servicios de Yahoo!. Algunos de los anuncios que te mostramos reflejan estos intereses. Puedes cancelar la publicidad y el contenido basados en intereses, la evaluación de contenido de comunicaciones con propósitos publicitarios, la divulgación de tu información a socios para equiparar y agregar información, y la recopilación de datos de análisis a través de sitios de socios que en su totalidad usan las herramientas de esta página. Tus opciones dentro de Ad Interest Manager no afectan a los intereses que específicamente indicas a Yahoo! en otras páginas. Más información.

La opción de exclusión de Flurry from Yahoo se gestiona por separado.

Tus categorías de interés ?

Usamos información sobre muchas de las páginas que has visitado, los anuncios que has visto y en los que has hecho clic, y algunas de tus búsquedas en Yahoo! para crear categorías de interés que nos ayudan a elegir el tipo de anuncios y contenido que verás. Puedes editar o deseleccionar las categorías aquí o cancelar en su totalidad los anuncios, el contenido y otras funciones que se basan en intereses. [Ver todas las categorías estándar](#)

Categoría de interés:

Este navegador no permite calificar anuncios basados en intereses. Asegúrate que has optado por la personalización de publicidad para poder ver las categorías de interés que este navegador puede calificar.

Publicidad personalizada:

Desactivada en este momento.

[Incluir](#)

Para que se aplique tu elección de optar por no usar anuncios en cada ordenador que usas, ingresa a tu cuenta Yahoo! y elige la opción persistente. Más información.

Tus actividades ?

Realizamos un resumen de muchas de tus actividades en Yahoo! aquí. Estas actividades ayudan a informar a nuestras categorías de interés y pueden ser utilizadas para otros tipos de publicidad personalizada. El resumen no se puede editar, pero en esta página puedes cancelar la publicidad personalizada en su totalidad.

Categoría que buscas:

Ninguna historia de categoría de búsqueda está disponible para este navegador.

Páginas y temas que visitas:

Ninguna página ni historia de temas está disponible para este navegador

Tu ordenador y las cookies ?

Podemos personalizar algunos anuncios basándonos en la información que nos envía tu ordenador y las cookies. Tu cancelación voluntaria de suscripción no es aplicable al uso de esta información demográfica o geográfica de los destinatarios. Así es como te vemos.

Ubicación:	Desconocido
Direcciones IP:	81.35.87.19
Sistema operativo:	unknown
Navegador:	
Resolución de pantalla:	1600x900
Profundidad de color:	24
Rango de edad:	Desconocido
Género:	Desconocido

[Política de Privacidad](#) | [Condiciones del Servicio](#)

Figura 4. Ad Interest Manager de Yahoo.

5.3. Exclusión de la publicidad personalizada

El objetivo fundamental de la publicidad personalizada de Yahoo es hacer que sus anuncios sean más pertinentes y útiles para cada usuario. No obstante, conscientes de que no todos los usuarios pueden estar de acuerdo con este sistema, Yahoo ha habilitado una opción donde puedes desactivar la publicidad y el contenido basados en intereses, la exploración del contenido de las comunicaciones con el propósito de enviar publicidad al usuario, y la recepción de datos de sitios asociados para sus productos de análisis. Este sistema se conoce con el nombre de "Opt-Out".

Los usuarios pueden optar por no recibir publicidad basada en intereses. Si lo hacen, entonces los sistemas automatizados de Yahoo no explorarán o analizarán el contenido de las comunicaciones enviadas y recibidas desde sus cuentas (incluyendo Correo Yahoo y contenido Messenger) para tales fines.

La desactivación de la publicidad personalizada se puede realizar tanto para los usuarios que tienen cuenta Yahoo como para los que no la tienen. La diferencia entre ambos casos radicará en la persistencia de las desactivaciones. Para que la desactivación sea tanto persistente como portátil, el usuario debe en su cuenta Yahoo y realizar la desactivación durante una sesión, de este modo, esta preferencia quedará almacenada en su cuenta y automáticamente se configurará en cualquier navegador que el usuario utilice, cada vez que ingrese a su cuenta. Sin embargo, para los usuarios que no tienen cuenta Yahoo (cabe mencionar que no necesitas tener una cuenta Yahoo para desactivar la publicidad basada en tus intereses) o aquellos usuarios que no hayan iniciado sesión en su cuenta, la desactivación solo será aplicable a un navegador específico en lugar de al usuario. En este último caso, la desactivación no es permanente.



Figura 5. Opción Opt-Out del servicio de publicidad basada en intereses de Yahoo.

Es importante añadir que la opción "Opt-Out" (Figura 5) se aplica sólo a la red de publicidad de Yahoo. No obstante, Yahoo ofrece anuncios tanto en Yahoo como en muchos otros sitios web. Del mismo modo, otras redes de anuncios también puede publicar anuncios en Yahoo, pero estas preferencias sobre el sistema "Opt-Out" y las categorías de interés no se aplican a estas otras redes de anuncios. Para

ver qué redes de anuncios pueden publicar anuncios en Yahoo se puede visitar la página sobre *Third Party and Affiliate Cookies on Yahoo* [22]. También se pueden ver muchas otras opciones de exclusión visitando la *NAI page* [24].

Hay, sin embargo, algunas categorías de interés personalizadas que no aparecen en la herramienta Ad Interest Manager, como por ejemplo las creadas específicamente para algunos anunciantes, como ya hemos visto antes. En este caso, Yahoo muestra las categorías de búsquedas y páginas vistas y los temas de Yahoo que podrían ser utilizados para el anuncio de personalización y, actualmente, no se puede optar por eliminar estas categorías. Si desea asegurarse de que está rechazando todo tipo de personalización basada en intereses, usted debe optar por usar el sistema "Opt-Out" en la página.

Cabe mencionar que Yahoo también usa señalizaciones web para una serie de propósitos, entre ellos, investigación y emisión de informes para algunos de nuestros clientes. Si optas por desactivar la publicidad basada en tus intereses, no usaremos la información recolectada a través de nuestra señalización web para fines de investigación e informes.

A pesar de todo, optar por no recibir publicidad basada en intereses a través del Ad Interest Manager no eliminará los intereses que se hayan seleccionado dentro de las herramientas de gestión de interés (donde estén disponibles) en Yahoo. Independientemente de sus opciones dentro del Interest Ad Manager, sus intereses declarados continuarán informando sobre el contenido y los anuncios que se muestran. Si el usuario no desea recibir anuncios o contenidos basados en sus intereses declarados, basta con quitar todos los intereses de sus herramientas de gestión de intereses declarados.

5.4. Privacidad de los usuarios de Yahoo

Como hemos visto, Yahoo utiliza la información acerca de muchas de las páginas que los usuarios han visitado, anuncios que han visto y sobre los que han hecho clic, algunas de sus búsquedas en Yahoo, y el uso de sus servicios para crear categorías de interés que les ayudan a elegir los tipos de anuncios y el contenido que cada usuario en concreto verá. Los usuarios pueden modificar o anular la selección de las categorías o inhabilitar los anuncios basados en intereses, el contenido y otras características.

La información recabada por Yahoo entra en los sistemas algorítmicos que ayudan a hacer una conjetura acerca de los intereses de sus usuarios.

Llegados a este punto, es un hecho a tener en cuenta que, de esta manera, podemos encontrar que algunos de los usuarios pueden advertir ciertos problemas sobre su privacidad.

A continuación, veremos algunos puntos importantes sobre el manejo de la información conseguida por Yahoo que tienen que ver con la privacidad de los usuarios.

5.4.1. Menores de 13 años

Los sistemas de publicación de anuncios de Yahoo no utilizan información sobre los visitantes que saben son menores de 13 años para dirigir la publicidad basada en intereses. Tampoco crean categorías dirigidas a los niños bajo la edad de 13 años.

5.4.2. Información sobre la salud

La salud es una categoría amplia con muchas subcategorías. Dado que todas las categorías de salud no son las mismas, Yahoo examina cada una individualmente. Algunos tipos de condiciones de salud, como las enfermedades de transmisión sexual, VIH / SIDA, el cáncer y los trastornos de salud mental, no tienen categorías de interés representadas en los sistemas de publicidad de Yahoo. Otros, como el resfriado y la gripe, problemas dentales, ardor de estómago y las alergias, sí que tienen categorías representadas.

Si un usuario quisiera recibir información basada en las categorías relacionadas con la salud, puede seleccionar dichos intereses en sus herramientas de gestión de intereses. Esta declaración es una señal para el equipo de Yahoo de que dicho usuario está interesado de forma explícita en esa categoría como tema de su contenido y la publicidad.

5.4.3. Obtención de datos demográficos

En muchos casos, la edad y el género reflejan los datos demográficos generales de la última persona que se registró en un equipo. Si el equipo es utilizado por varios miembros de la familia o si el usuario está en un lugar público, no se almacena ni se muestra esta información.

Yahoo obtiene información sobre la ubicación del usuario a través de diversos métodos, tales como los datos que el usuario indica cuando se registra en una cuenta de Yahoo o cuando entra en una ubicación en Yahoo Maps.

Yahoo utiliza esta información para personalizar los anuncios que aparecen tanto en Yahoo como otros sitios web. Para obtener más información acerca de

cómo Yahoo obtiene información sobre su ubicación, Yahoo dispone de una página especializada: *Yahoo Location privacy page* [25].

5.4.4. *Compartición de información*

Yahoo no comparte las categorías de intereses etiquetadas a los usuarios con los anunciantes u otros sitios web, pero se podría dar el caso en el que, por ejemplo, una compañía aérea podría querer mostrar anuncios de los destinos de vacaciones para los usuarios de Yahoo que han mostrado cierto interés en esos destinos. El anunciante proporciona información acerca del público que les gustaría alcanzar y Yahoo muestra sus anuncios a través de su red de anuncios a las personas que responden a las características que los anunciantes han dado.

Tampoco se comparten listas de usuarios con anunciantes u otros sitios web cuando un anunciante trabaja con Yahoo en una campaña publicitaria personalizada.

5.4.5. *Inhabilitación de la publicidad*

En Yahoo no es posible inhabilitar los anuncios por completo.

Como la mayoría de sitios web, Yahoo se sostiene con la publicidad. La mayoría de los productos y servicios que se ofrecen, tales como el correo electrónico, intercambio de fotos en Flickr, Yahoo Noticias y muchos otros servicios son, en gran parte, libres de cargo para los usuarios porque muestran publicidad.

5.4.6. *Cookies*

Para la recogida de información Yahoo utiliza cookies, que tienen una cierta caducidad.

La política de Yahoo dice que una "Opt-Out cookie" expira a los 20 años desde la fecha en que se creó. El resto de las cookies de Yahoo deben caducar dos años después de que se establecen. Yahoo utiliza las cookies del navegador para la personalización de la publicidad. Yahoo no utiliza "cookies de Flash" u otros "objetos locales compartidos" para la publicidad basada en intereses.

Por el contrario, las categorías de interés almacenadas y los datos utilizados para crearlos se descartan mucho antes. Esto se debe a que las categorías de interés son un promedio móvil de los intereses de una persona. Esto significa que cambian con el tiempo en función de la actividad más reciente de una persona en Yahoo. La mayoría de las categorías tienen una vida útil desde unos pocos días hasta varios meses. En Yahoo creen que es más probable que la

gente esté interesada en los anuncios que reflejen las visitas más recientes y la actividad.

6. Planificación

En este apartado describiremos la planificación del trabajo, no solo del desarrollo de la herramienta, que es lo que más tiempo y esfuerzo ha supuesto, sino de todas las tareas realizadas durante el desarrollo del proyecto.

Consta de cuatro fases. Cada una de ellas comprende tareas específicas que, en su conjunto, componen el proyecto:

1. Planteamiento del proyecto.
 - a. Definición de la primera versión de la herramienta.
 - b. Análisis de las propuestas de software para el desarrollo.
 - c. Recopilación de información sobre la publicidad de Yahoo.
2. Desarrollo de la herramienta.
 - a. Elección de software.
 - b. Lectura de información sobre software.
 - c. Familiarización con el software: primeras pruebas.
 - d. Programación con iMacros.
 - e. Programación con Javascript.
 - f. Programación con Shell.
 - g. Programación con MySQL.
 - h. Implementación de mejoras.
3. Experimentos.
 - a. Instalación de la máquina virtual.
 - b. Instalación de software necesario.
 - c. Definición de experimentos.
 - d. Ejecución de experimentos.
 - e. Obtención de resultados.
 - f. Redefinición de experimentos.
 - g. Ejecución de experimentos.
 - h. Obtención y análisis de los resultados.
4. Memoria.
 - a. Planteamiento.
 - b. Redacción.

La planificación está representada en un diagrama de Gantt, disponible en el ANEXO V, en el que se detalla el tiempo que se ha dedicado a cada una de las tareas.

7. Desarrollo de la herramienta

Tal y como adelantamos en el apartado 3, el objetivo de este proyecto es realizar un estudio de la publicidad basada en intereses en el caso concreto de Yahoo. Para ello, ha sido necesario desarrollar una herramienta que interactúa de forma automática con distintos servicios Yahoo (buscador, correo, Flickr y Screen).

La herramienta es totalmente configurable (ofrece la posibilidad de elegir el servicio, los tiempos de interacción, el número de repeticiones y los términos con los que se realizan las pruebas) y, mediante los distintos experimentos llevados a cabo, proporcionará datos relevantes para el análisis del comportamiento del sistema que utiliza Yahoo, y así basar la publicidad en los intereses de sus usuarios, personalizada para cada uno de ellos.

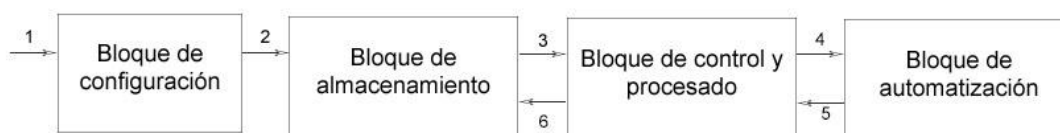


Figura 6. Estructura general de la herramienta.

Tal como muestra la Figura 6, la herramienta consta de cuatro bloques (almacenamiento, configuración, control y procesado y automatización), por los que se moverá durante el desarrollo de los experimentos.

1. Bloque de configuración: obtiene las opciones de definen el experimento.
2. Bloque de almacenamiento: extrae de la base de datos la información necesaria para realizar el experimento.
3. Bloque de control y procesado: procesa los datos adquiridos antes de entrar al siguiente bloque. Se encarga de controlar el funcionamiento general de la herramienta.
4. Bloque de automatización: automatiza el navegador para realizar el experimento definido y obtiene información relevante.
5. Bloque de control y procesado (2): procesa los datos obtenidos de la automatización.
6. Bloque de almacenamiento (2): almacena los resultados obtenidos una vez procesados.

Para llevar a cabo el desarrollo de la herramienta, se han empleado tres tecnologías, cada una de las cuales se centra en una parte específica de la herramienta. La combinación de las tres conforma así la herramienta automática y flexible que necesitamos para realizar el estudio.

El bloque de configuración es sencillamente un documento en el que se configuran las distintas opciones antes de lanzar los experimentos.

Aunque lo explicaremos más en profundidad en el apartado 7.1, podemos adelantar que para automatizar la interacción con los productos Yahoo (bloque de automatización), utilizaremos la herramienta iMacros. Para el almacenaje y control de los datos (bloque de almacenamiento) utilizaremos una base de datos MySQL. Y por último, para la gestión general de los experimentos y administración de los datos, así como el control general de la herramienta (bloque de control y procesado), utilizaremos un intérprete de comandos, codificado mediante Shell Scripting.

A lo largo de este apartado se profundizará en el funcionamiento de esta herramienta a través de las tecnologías utilizadas (apartado 7.1). No obstante, en el ANEXO I, ANEXO II y ANEXO III encontramos información relevante acerca de la programación de la herramienta.

7.1. Tecnologías utilizadas

7.1.1. *Bloque de automatización: Automatización de las búsquedas con iMacros*

Para el desarrollo de este bloque, es necesaria una herramienta capaz de automatizar procesos en el navegador web.

QTP, Rational Functional Tester (RFT), Selenium, Sahi y Webdrive son algunos ejemplos de las muchas herramientas comerciales e infraestructuras de código abierto para las pruebas automatizadas de aplicaciones web. Estas herramientas suelen permitir grabar y reproducir acciones en todos los navegadores y proporcionan controladores de lenguaje para escribir scripts de prueba.

iMacros [26] es la tecnología escogida para la automatización del navegador, que es el encargado de interactuar con Yahoo, efectuando las tareas pertinentes como realizar búsquedas en el buscador, Flickr o Screen o enviando correos electrónicos de una cuenta a otra.

Aunque la acción de enviar un correo no es igual que hacer una búsqueda, por simplicidad a la hora de explicar el funcionamiento de la herramienta, llamaremos "búsqueda" a la interacción con cualquiera de los cuatro productos.

Cabe mencionar que iMacros no dispone de controladores de lenguaje, así que desarrollaremos esa necesidad creando nuestro propio script controlador en lenguaje Javascript.

7.1.1.1. ¿Qué es iMacros?

iMacros es un complemento para navegadores, diseñado para automatizar tareas repetitivas en la web. Es una herramienta que graba acciones realizadas sobre un navegador, en este caso el navegador escogido es Firefox, y las convierte en código que posteriormente podemos utilizar para repetirlas de forma periódica, lo que nos es realmente útil para realizar nuestros experimentos.

En otras palabras, iMacros es una extensión del navegador web que permite grabar, reproducir y almacenar las interacciones que llevamos a cabo con la web. Además es capaz de generar estadísticas y capturar todo o parte del contenido de una determinada página web. Esta herramienta también se puede configurar de forma que se puedan variar los datos de entrada e incluso se pueden crear scripts, tal y como hemos procedido en este caso, para que realice diversas funcionalidades [27].

7.1.1.2. Utilización de iMacros en la herramienta

Hemos generado un código Javascript que, a partir de la definición de un experimento, es capaz de controlar el lanzamiento de distintas iMacros de forma que se interactúe con Yahoo y se lleve a cabo el experimento definido, obteniendo los datos pertinentes para realizar el posterior estudio.

7.1.2. Bloque de control y procesado: Shell Scripting

Para el diseño de este bloque necesitaremos un intérprete de comandos.

Un intérprete de comandos es un programa informático que con capacidad para traducir las órdenes que introducen los usuarios, mediante un conjunto de instrucciones [28].

La herramienta se ha desarrollado sobre Linux, ya que este sistema operativo es más flexible que Windows a la hora de procesar datos y realizar análisis, por lo que el Shell Scripting será el encargado de gestionar el conjunto de la herramienta y procesar los datos obtenidos durante los experimentos.

Para ver cómo lo lleva a cabo, explicaremos más en detalle su funcionamiento.

7.1.2.1. ¿Qué es un Shell?

Shell es un intérprete de comandos, encargado de traducir las órdenes del usuario a comandos que sólo entiende el sistema operativo. De esta forma, gestionaremos el funcionamiento de nuestra herramienta mediante los comandos que consideremos oportunos.

7.1.2.2. Shell Scripting

Shell no es, estrictamente hablando, un lenguaje de programación pero, si se construye un guión de comandos en un fichero ejecutable, el intérprete traducirá al sistema operativo las órdenes línea a línea para que éste las lleve a cabo. Esta forma de programación es lo que se conoce como Shell Scripting y es lo que utilizamos en la construcción de esta herramienta [29].

7.1.2.3. Utilización de Shell Scripting en la herramienta

Este código es el encargado de gestionar el funcionamiento y el estado de la herramienta. A través de él accedemos a las distintas tecnologías y nos movemos de un bloque a otro de la herramienta (figura 5). Gestiona, además, la liberación de recursos, el procesamiento de los documentos con la información extraída y la salida y los mensajes de error que emite el programa.

7.1.3. Bloque de almacenamiento: Gestión de la información con MySQL

Para el almacenaje de datos y la gestión de la información generada, necesitaremos una base de datos.

Una base de datos es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso [30].

Existen varios tipos de bases de datos, que se clasifican según la variabilidad de los datos o del contenido. En nuestro caso almacenaremos la información en distintas tablas de una base de datos que se desarrolla en MySQL.

7.1.3.1. Bases de datos en MySQL

Una base de datos es una colección de datos organizada para permitir el acceso a dichos datos de una forma relativamente fácil. Es por esto que, para manejar la información extraída del Ad Interest Manager de Yahoo, utilizaremos una base de datos.

Para este trabajo utilizaremos, en concreto, una base de datos relacional programada con SQL (Structured Query Language), que es un lenguaje estándar para especificar acceso, creación y modificación de bases de datos relacionales, como es MySQL.

MySQL es un sistema de gestión de bases de datos relacional, multihilo y multiusuario, desarrollado en su mayor parte en ANSI C (estándar publicado por el Instituto Nacional Estadounidense de Estándares, para el lenguaje de programación C) [31].

Como ya se ha mencionado, MySQL es un sistema de administración relacional de bases de datos. Una base de datos relacional archiva datos en tablas separadas, lo que permite mayor velocidad y flexibilidad. Las tablas de una misma base de datos están conectadas, de forma que puedan combinarse datos de diferentes tablas.

MySQL es un software de fuente abierta, lo que significa que cualquier persona puede usarlo y modificarlo, es decir, cualquier persona puede bajar el código fuente de MySQL y ajustarlo a sus necesidades sin coste alguno.

7.1.3.2. Utilización de MySQL en la herramienta

La herramienta dispone de una base de datos a la que accederá tanto para la extracción de información necesaria para lanzar un experimento como para almacenar, en distintas tablas, la información obtenida de los distintos experimentos una vez procesada.

Después de explicar qué tecnologías hemos utilizado para la implementación de cada uno de los bloques que componen la herramienta, pasaremos a describir el funcionamiento y desarrollo de cada uno de estos módulos en los siguientes apartados.

7.2. Bloque de configuración

En la introducción de este apartado 7.1 ya hemos adelantado que este bloque no va mucho más allá de un documento de texto donde se configuran las opciones que definen cada uno de los experimentos.

El fichero de configuración consta de cuatro campos a rellenar antes de lanzar un experimento:

1. **Número de repeticiones:** búsquedas que se realizarán de cada término antes de cerrar el navegador.
2. **Duración estimada total:** tiempo en segundos que se desea emplear en realizar todas las búsquedas. Se divide entre el número de repeticiones para obtener cada cuántos segundos se inicia una nueva búsqueda. No es el tiempo real del experimento porque no se tienen en cuenta los tiempos de carga y procesamiento del navegador.
3. **Producto:** producto Yahoo sobre el que se realiza la búsqueda. Cuatro configurados: buscador de Yahoo, Flickr, Yahoo Screen y servicio de correo electrónico ofrecido por Yahoo.

4. **Términos:** términos a buscar en función del número seleccionado. Se extraen de la base de datos. (Ver apartado 7.3. para más información sobre los términos almacenados).

Una vez definido el experimento, se pasa al bloque de almacenamiento, que obtendrá de la base de datos los datos correspondientes a la selección establecida en el cuarto campo.

En el ANEXO IV encontramos un ejemplo del fichero de configuración.

7.3. Bloque de almacenamiento: Acceso 1

El módulo de almacenamiento consta de una base de datos, en la que se almacenarán tres tablas:

1. Tabla de categorías (Tabla 1 en la Figura 7): en ella se almacenan las distintas categorías en las que los usuarios de Yahoo pueden ser etiquetados.
2. Tabla general (Tabla 2 en la Figura 7): en ella se almacenan los resultados generales de los experimentos.
3. Tabla específica (Tabla 3 en la Figura 7): en ella se almacenan los resultados pormenorizados de cada una de las búsquedas de un experimento.

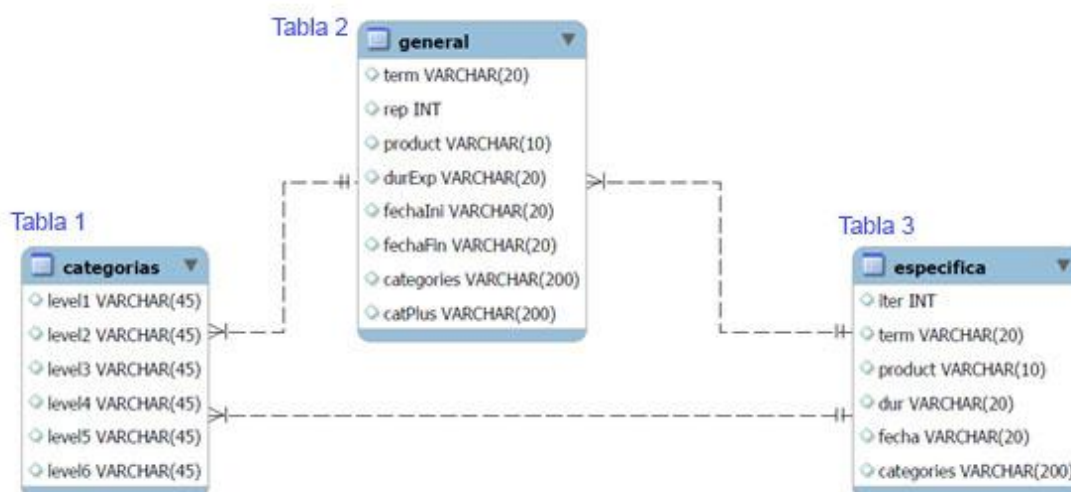


Figura 7. Estructura de la base de datos.

Las relaciones entre las tablas las definiremos de la siguiente manera:

- Los términos almacenados en la *tabla de categorías* serán los que utilice la herramienta para interactuar con los servicios Yahoo, de este modo, serán

los términos que se almacenen en el campo "term" de las tablas *general* y *específica*.

- La tabla específica detalla los resultados de la *tabla general*. En ella se genera una nueva entrada por cada una de las repeticiones y/o categorías etiquetadas en los experimentos de la *tabla general*.

A continuación explicaremos más en detalle qué se almacena en cada una de las tablas de la base de datos ayudándonos de ejemplos.

Tabla de categorías

Como podemos apreciar en la Figura 7, esta tabla dispone de seis campos. En ellos se almacenan las distintas categorías de intereses de Yahoo, que sirven a su vez como los términos empleador para la interacción de la herramienta con los distintos servicios.

Yahoo dispone de un listado de categorías sobre las que los usuarios pueden estar interesados y que son en sí mismas lo que nosotros llamamos etiquetas. Así, en función de las acciones de los usuarios en su sitio web, se determinan sus intereses y se les asignan determinadas categorías. El listado de categorías puede leerse en la página correspondiente [32].

Para su almacenaje en la base de datos, hemos estructurado las categorías en una jerarquía de niveles que nos ayudará al almacenaje en la *tabla de categorías*. Vemos un ejemplo de ello:

Viajes, Destinos, Europa, Turquía
Viajes, Destinos, Europa, UK
Viajes, Destinos, Europa, UK, Gales
Viajes, Destinos, Europa, UK, Inglaterra
Viajes, Destinos, Europa, UK, Inglaterra, Escocia
Viajes, Destinos, Europa, UK, Inglaterra, Londres

En estas líneas vemos ejemplos de la categoría "Viajes" y algunas de sus subcategorías. Este ejemplo nos servirá para entender cómo hemos estructurado la jerarquía y su almacenamiento en la base de datos.

La categoría "Viajes" pertenecerá al nivel 1 (level1) y estará almacenada en la primera columna de la *tabla de categorías* de nuestra base de datos. Del mismo modo, 'Destinos' estaría en la segunda columna, 'Europa' en la tercera y así sucesivamente. Es decir, estas categorías estarían almacenadas en la tabla de la base de datos de la siguiente forma:

level1	level2	level3	level4	level5	level6
Viajes	Destinos	Europa	Turquía	NULL	NULL
Viajes	Destinos	Europa	UK	NULL	NULL
Viajes	Destinos	Europa	UK	Gales	NULL
Viajes	Destinos	Europa	UK	Inglaterra	NULL
Viajes	Destinos	Europa	UK	Inglaterra	Escocia
Viajes	Destinos	Europa	UK	Inglaterra	Londres

Tabla 4. Ejemplo de almacenamiento de la *tabla de categorías* en la base de datos.

Con esta estructura aplicada a todas las categorías almacenadas, la selección de nivel en la definición de los experimentos conlleva la elección de los términos concretos que vamos a buscar. Es decir, la utilización del nivel 1 implicará la utilización de términos muy generales, que se van haciendo más específicos conforme seleccionamos niveles superiores.

Tabla general

Como podemos apreciar en la Figura 7, esta tabla dispone de ocho campos.

Esta tabla será la encargada de recoger los datos generales en cada experimento. Para ello almacenamos lo siguiente:

- "term": Se refiere al término de búsqueda en cada ocasión.
- "rep": Se refiere al número de búsquedas totales que se han realizado de cada término en un experimento.
- "product": Se refiere al producto Yahoo sobre el que realizamos la búsqueda (buscador, Screen, Flickr o correo de Yahoo).
- "durExp": Se refiere al tiempo estimado en segundos que se dedica a todas las búsquedas de un mismo término del experimento.
- "fechaIni": Se corresponde con la fecha y hora de inicio de la búsqueda en cuestión.
- "fechaFin": Se corresponde con la fecha y hora de fin de la búsqueda en cuestión.
- "categories": Almacena todas las categorías que se etiquetan para cada término del experimento.
- "catPlus": Almacena la(s) categoría(s) más etiquetadas para cada término del experimento.

Un ejemplo de almacenamiento de esta tabla sería el siguiente:

term	rep	product	durExp	fechaIni	fechaFin	categories	catPlus
Gales	3	buscador	120	02/04/2015 12:17	02/04/2015 12:19		
Inglaterra	3	buscador	120	02/04/2015 12:19	02/04/2015 12:22	Viajes & Viajes > Destinos >Europa	Viajes
Tenis	3	buscador	120	02/04/2015 12:22	02/04/2015 12:06	Deportes	Deportes

Tabla 5. Ejemplo de almacenamiento de la *tabla general* en la base de datos.**Tabla específica**

Como podemos apreciar en la Figura 7, esta tabla dispone de seis campos.

Esta tabla será la encargada de recoger los datos específicos correspondientes a cada iteración de cada búsqueda realizada en cada uno de los experimentos. Para ello almacenamos lo siguiente:

- "iter": Se corresponde con la el número de búsqueda de un término. Es decir, si se lanza un experimento en el que se busca cada término tres veces, en "iter" se almacenarán los números 1, 2 y 3 correspondientes a cada una de las búsquedas para un mismo término.
- "term": Se refiere al término de búsqueda en cada ocasión.
- "product": Se refiere al producto Yahoo sobre el que realizamos la búsqueda (buscador, Screen, Flickr o correo de Yahoo).
- "dur": Se refiere al tiempo estimado en segundos que se dedicará a cada búsqueda de cada término.
- "fecha": Se corresponde con la fecha y hora de inicio de la búsqueda en cuestión.
- "categories": Almacena de una en una las categorías que se etiquetan en cada iteración.

Un ejemplo de almacenamiento de esta tabla sería el siguiente:

iter	term	product	dur	fecha	categories
1	Gales	buscador	40	02/04/2015 12:17	
2	Gales	buscador	40	02/04/2015 12:17	
3	Gales	buscador	40	02/04/2015 12:18	
1	Inglaterra	buscador	40	02/04/2015 12:19	
2	Inglaterra	buscador	40	02/04/2015 12:19	Viajes
3	Inglaterra	buscador	40	02/04/2015 12:20	Viajes
3	Inglaterra	buscador	40	02/04/2015 12:21	Viajes > Destinos > Europa
1	Tenis	buscador	40	02/04/2015 12:22	Deportes
2	Tenis	buscador	40	02/04/2015 12:23	Deportes
3	Tenis	buscador	40	02/04/2015 12:24	Deportes

Tabla 6. Ejemplo de almacenamiento de la *tabla específica* de la base de datos.

El primer acceso al bloque de almacenamiento (paso 2 en la Figura 6) se emplea para acceder a la base de datos y, sabiendo del módulo de configuración qué términos se van a emplear para realizar las búsquedas en el experimento, descargar el nivel correspondiente de la *tabla de categorías* (representada en la Figura 7).

Con esto, la herramienta dispone de todos los términos que utilizará durante el experimento y puede pasar al bloque de control y procesado.

7.4. Bloque de control y procesado: Acceso 1

En este primer acceso, la herramienta utiliza este bloque para varias cosas:

- Controlar la salida estándar y los mensajes de error que se producen.
- Organizar los datos extraídos de la base de datos: selección de los términos de búsqueda.
- Creación de una carpeta de *backup*.
- Lanzar el bloque de automatización.
- Liberar recursos.

Como este bloque administra el funcionamiento general de la herramienta, hay que tener en cuenta que antes de iniciar el experimento, tenemos que borrar cualquier documento que exista correspondiente a experimentos anteriores para que no se interfieran con los que van a crearse, así que este bloque realiza dicho borrado.

En cada experimento se lleva a cabo la búsqueda de una sucesión de términos, que han sido extraídos de la base de datos y, tras ser procesados, se guardan en un fichero. De este fichero se irán leyendo uno a uno los términos para realizar las diferentes búsquedas dentro del bloque de automatización.

La creación de la carpeta de *backup* es lo que su propio nombre indica: el bloque se encarga de crear una carpeta donde almacenará todos los resultados obtenidos para tener una copia local aparte de la copia en la base de datos.

Llegados a este punto, ahora tenemos que lanzar el archivo Javascript en el navegador para empezar las búsquedas.

El mecanismo a realizar será el siguiente:

- Selección del término a buscar en cada ocasión.
- Apertura de una ventana de Firefox y ejecución del Javascript que controla las iMacros.
- Comprobación periódica de si hay alguna iMacro ejecutándose o el Javascript ha llegado al final. De ser así, se iniciará una nueva búsqueda en un nuevo navegador.

Para evitar el malgasto de RAM, las ventanas de Firefox se cerrarán cada cinco términos, es decir, mantendremos las ventanas abiertas (aunque sin hacer nada) hasta que tengamos cinco, en ese caso mataremos los procesos de Firefox y liberaremos recursos llamando a un nuevo ejecutable Shell, cuya única función es, como decimos, liberar la memoria caché.

7.5. Bloque de automatización

Como se ha mencionado anteriormente, se ha programado un archivo Javascript como encargado de controlar las iMacros. Así que, llegados a este punto, el bloque de control y procesado se encarga de invocar este archivo Javascript dentro del navegador y éste será el encargado de llamar y controlar las iMacros de forma ordenada para reproducir el experimento definido en el bloque de configuración.

Por lo tanto, el objetivo general de esta parte de la herramienta será:

- Leer opciones marcadas por el usuario en el fichero de configuración.
- Realizar las búsquedas correspondientes.
- Comprobar las categorías etiquetadas por Yahoo tras cada búsqueda.
- Extraer, en caso de que existan, las distintas categorías etiquetadas.

No obstante, a fin de realizar un trabajo exhaustivo y preciso para llevar a cabo el correcto funcionamiento de toda la herramienta, el archivo Javascript realizará otras funciones secundarias que no participan activamente del objetivo general, pero que son igualmente necesarias para, como decíamos, el funcionamiento global de la herramienta diseñada. En el ANEXO II se listan las actividades que se han programado en este bloque.

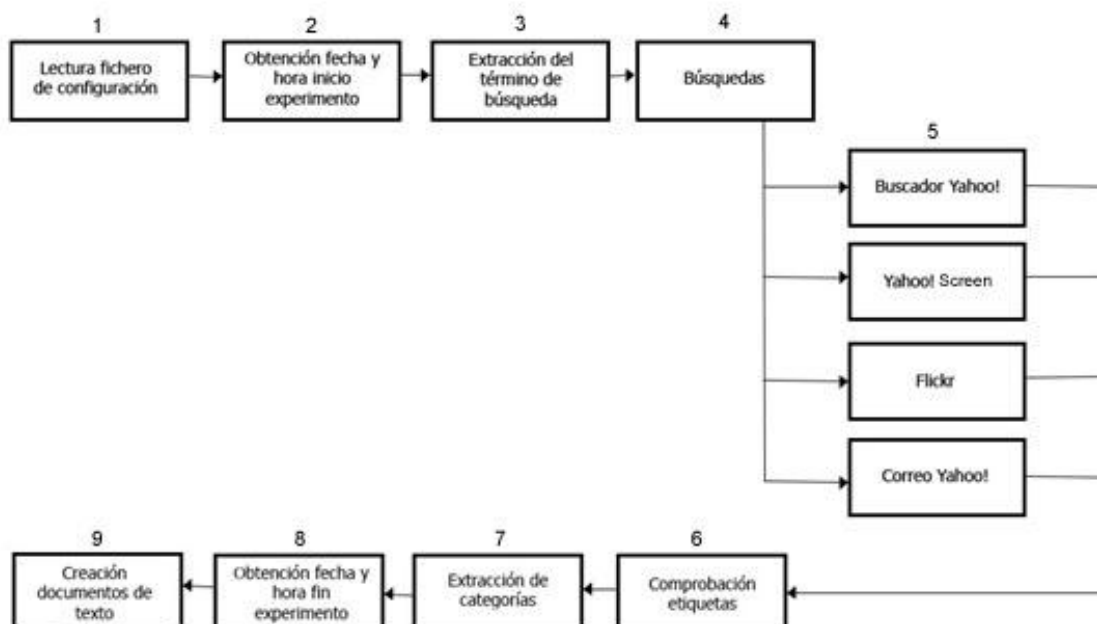


Figura 8. Esquema del bloque de automatización.

A continuación llevaremos a cabo la descripción paso a paso de este bloque.

1. Lectura del archivo de configuración.

No es estrictamente una lectura, pues los datos ya los ha extraído el bloque de control y procesado, pero ahora se tienen acceso a ellos para determinar las características del experimento definido.

2. Obtención de la fecha y la hora al Inicio del experimento.

Necesitamos conocer la fecha y hora de inicio del experimento para almacenarlas en la base de datos.

3. Extracción del término de búsqueda.

Sencillamente, la herramienta se encarga de llevar un control de qué término se busca en cada ocasión.

4. Búsquedas.

En esta parte, la herramienta ejecuta las órdenes precisas para realizar las búsquedas y posteriormente extraer los diferentes datos.

Recordemos que el bloque de control y procesado ha sido el encargado de invocar el Javascript que controla este bloque, de forma que este organiza y lanza las distintas iMacros.

5. Selección del producto Yahoo.

La iMacros que llaman al buscador, Flickr y Screen son similares en funcionamiento. La que llama al servicio de correo tiene características similares, ya que no es lo mismo automatizar una búsqueda que enviar un correo electrónico.

6. Comprobar si existen categorías etiquetadas

Para realizar esta comprobación, tenemos que dirigirnos a la página en la que se muestran las categorías etiquetadas, es decir, la página del Ad Interest Manager (apartado 5.2) y, una vez en esta página, lanzar la iMacro que comprueba si hay etiquetas.

La Figura 9 muestra el aspecto de la página en que se etiquetan las categorías cuando no hay ninguna etiquetada.

La Figura 10 muestra el aspecto de la misma página cuando tenemos alguna categoría etiquetada.

Tus categorías de interés

Usamos información sobre muchas de las páginas que has visitado, los anuncios que has visto y en los que has hecho clic, y algunas de tus búsquedas en Yahoo! para crear categorías de interés que nos ayudan a elegir el tipo de anuncios y contenido que verás. Puedes editar o deseleccionar las categorías aquí o cancelar en su totalidad los anuncios, el contenido y otras funciones que se basan en intereses. [Ver todas las categorías estándar](#)

Categoría de interés:

Este navegador no permite calificar anuncios basados en intereses. Asegúrate que has optado por la personalización de publicidad para poder ver las categorías de interés que este navegador puede calificar.

Publicidad personalizada:
Activada en este momento.

[Excluir](#)

Debes permitir las cookies de Yahoo! para optar por excluir la publicidad personalizada. Para que tu opción de excluir la publicidad personalizada sea de aplicación en cada ordenador que utilices, debes acceder a tu cuenta Yahoo! [Más información.](#)

Figura 9. Ad Interest Manager sin categorías de interés.

Tus categorías de interés [?](#)

Usamos información sobre muchas de las páginas que has visitado, los anuncios que has visto y en los que has hecho clic, y algunas de tus búsquedas en Yahoo! para crear categorías de interés que nos ayudan a elegir el tipo de anuncios y contenido que verás. Puedes editar o deseleccionar las categorías aquí o cancelar en su totalidad los anuncios, el contenido y otras funciones que se basan en intereses. [Ver todas las categorías estándar](#)

Categoría de interés: Configurar:

Entretenimiento	<input type="button" value="ACTIVAR"/> <input type="button" value="DESACTIVAR"/>	<div> Publicidad personalizada: Activada en este momento. <input type="button" value="Excluir"/> </div> <p>Debes permitir las cookies de Yahoo! para optar por excluir la publicidad personalizada. Para que tu opción de excluir la publicidad personalizada sea de aplicación en cada ordenador que utilices, debes acceder a tu cuenta Yahoo! Más información.</p>
Tecnología	<input type="button" value="ACTIVAR"/> <input type="button" value="DESACTIVAR"/>	
Tecnología > Hardware	<input type="button" value="ACTIVAR"/> <input type="button" value="DESACTIVAR"/>	
Tecnología > Productos Electrónicos	<input type="button" value="ACTIVAR"/> <input type="button" value="DESACTIVAR"/>	
Tecnología > Productos Electrónicos > Móviles > Teléfonos Móviles	<input type="button" value="ACTIVAR"/> <input type="button" value="DESACTIVAR"/>	
<input type="button" value="Guardar"/>		

Figura 10. Ad Interest Manager con categorías de interés.

7. Extracción de información cuando sí hay categorías etiquetadas.

En este punto, la herramienta se encarga de la extracción de categorías etiquetadas, distinguiendo entre dos casos.

Si observamos la Figura 11, podemos ver que, para el navegador Firefox, cuando Yahoo nos etiqueta en muchas categorías, sólo ocho se muestran en la página y, debajo de ellas, aparece el desplegable "Mostrar todo", sobre el que deberemos pulsar si queremos ver todas las categorías que Yahoo nos ha asignado.

Es por esto que diferenciaremos entre dos casos a la hora de la extracción de datos: la extracción de las categorías que se muestran directamente sobre la página web (a las que llamaremos categorías visibles) y la extracción de las categorías accesibles desde el desplegable (a las que llamaremos categorías invisibles). Para ver la diferencia entre las extracciones, ver el código en el ANEXO III.

Tus categorías de interés ?

Usamos información sobre muchas de las páginas que has visitado, los anuncios que has visto y en los que has hecho clic, y algunas de tus búsquedas en Yahoo! para crear categorías de interés que nos ayudan a elegir el tipo de anuncios y contenido que verás. Puedes editar o deseleccionar las categorías aquí o cancelar en su totalidad los anuncios, el contenido y otras funciones que se basan en intereses. [Ver todas las categorías estándar](#)

Categoría de interés: Configurar:

Consumo > Comida y Nutrición	<input type="button" value="ACTIVAR"/> <input type="button" value="DESACTIVAR"/>	<div>Publicidad personalizada:</div> <div>Activada en este momento.</div> <div><input type="button" value="Excluir"/></div> <p><small>Debes permitir las cookies de Yahoo! para optar por excluir la publicidad personalizada. Para que tu opción de excluir la publicidad personalizada sea de aplicación en cada ordenador que utilices, debes acceder a tu cuenta Yahoo! Más información.</small></p>
Entretenimiento	<input type="button" value="ACTIVAR"/> <input type="button" value="DESACTIVAR"/>	
Entretenimiento > Música	<input type="button" value="ACTIVAR"/> <input type="button" value="DESACTIVAR"/>	
Tecnología	<input type="button" value="ACTIVAR"/> <input type="button" value="DESACTIVAR"/>	
Tecnología > Hardware	<input type="button" value="ACTIVAR"/> <input type="button" value="DESACTIVAR"/>	
Tecnología > Productos Electrónicos	<input type="button" value="ACTIVAR"/> <input type="button" value="DESACTIVAR"/>	
Tecnología > Productos Electrónicos > Móviles > Teléfonos Móviles	<input type="button" value="ACTIVAR"/> <input type="button" value="DESACTIVAR"/>	
Viajes	<input type="button" value="ACTIVAR"/> <input type="button" value="DESACTIVAR"/>	
<input type="button" value="Mostrar todo"/>	<input type="button" value="Guardar"/>	

Figura 11. Ad Interest Manager con más de ocho categorías de interés.

8. Obtención de la fecha de fin del experimento.

A la hora de almacenar los datos relevantes del experimento para su posterior estudio, necesitaremos guardar la fecha de fin del experimento, igual que ya almacenamos la de inicio al principio de la ejecución del código.

9. Concatenar datos para pasarlos al documento de texto con la información.

Una vez extraída toda la información que necesitamos para realizar el estudio, la recuperamos de todas las variables y la colocamos en un documento de texto de forma que resulte práctico tratarla después.

7.6. Bloque de control y procesado: Acceso 2

De vuelta en el bloque de control y procesado, la herramienta dispone de los archivos de texto con los resultados del experimento, procedentes del bloque de automatización. No obstante, antes de almacenar estos datos en la base de datos, es necesario procesar los documentos y adaptarlos a nuestras necesidades. De esto se encargará este bloque en este segundo acceso.

En este segundo acceso, por tanto, la herramienta utiliza este bloque para:

- Procesar y organizar los archivos devueltos por el bloque de automatización.
- Realizar la copia de *backup*.
- Acceder al bloque de almacenamiento.

Así pues, la herramienta procesa los datos de los documentos para generar nuevos documentos que permitan subir la información relevante a la base de datos y copiar después todos los documentos de texto a la carpeta de *backup* ya mencionada, que posteriormente se comprime en *zip* y que contiene todos los datos obtenidos en una copia local.

Se crearán dos documentos que se le pasarán al bloque de almacenamiento.

El primero de ellos, para la tabla 3 (Figura 7), almacenará los siguientes datos:

- Número de iteración. Como podemos realizar experimentos en los que se realiza la búsqueda de un término más de una vez, este número indicará cuál es dicha búsqueda de entre las que se harán de dicho término.
- Término de búsqueda.
- Producto Yahoo sobre el que se realiza la búsqueda.
- Duración de la búsqueda.
- Fecha y hora en la que se inicia la búsqueda.
- Categorías etiquetadas tras la búsqueda.

El segundo, para la tabla 2 (Figura 7), almacenará:

- Término de búsqueda.
- Número de repeticiones. Son las veces que se repite una búsqueda de un mismo término.

- Producto Yahoo sobre el que se realiza la búsqueda.
- Duración total estimada del experimento.
- Fecha y hora en la que se inicia el experimento.
- Fecha y hora en la que finaliza el experimento.
- Categorías etiquetadas en cada búsqueda.
- Categoría(s) más etiquetada(s) en todo el experimento.

Todos los datos que necesitamos son fácilmente accesibles desde el documento que nos proporciona iMacros excepto la(s) categoría(s) más etiquetada(s). Para obtener este campo tendremos que realizar algunas acciones sobre los datos que ya tenemos, que consisten en:

- Copiar todas las categorías etiquetadas en un documento de texto.
- Separarlas en una línea cada una.
- Contar cuántas veces aparece cada una de ellas.
- Quedarnos con la(s) que más aparece(n).

Una vez se obtienen los documentos tal y como necesitamos, se vuelve a acceder al bloque de almacenamiento.

7.7. Bloque de almacenamiento: Acceso 2

En este segundo acceso al bloque de almacenamiento, la herramienta ya ha realizado las búsquedas pertinentes en Yahoo, se han procesado los datos y se han almacenado en documentos de texto organizados.

Con esto, accedemos al bloque de almacenamiento para almacenar los datos recopilados en las tablas 2 y 3 (Figura 7) de la base de datos.

Para más información sobre cómo quedan almacenados estos datos, ver el apartado 7.3.

8. Batería de experimentos

Para obtener conclusiones sobre cómo funciona el sistema de publicidad basado en intereses de Yahoo, a partir de la herramienta desarrollada, hemos realizado una batería de experimentos que, correctamente interpretados, nos acercan a la idea de dicho funcionamiento.

La batería de experimentos se ha realizado utilizando cuatro de los servicios proporcionados por Yahoo, que son: Buscador Yahoo, Correo Yahoo, Flickr y Yahoo Screen. La Tabla 7 sirve como resumen de los distintos experimentos realizados para el estudio y que se explicarán más en profundidad a lo largo de este apartado.

De las pruebas ejecutadas en estos cuatro servicios hemos obtenido diferentes resultados, tanto positivos como negativos, que analizaremos en este apartado con el fin de entender, como ya hemos mencionado, cómo entendemos que puede funcionar la publicidad basada en intereses en Yahoo.

Servicio Yahoo	Experimentos			
	Número	Términos buscados	Tiempo entre búsquedas (s)	Repeticiones
Correo	1	5	180	10
	2	4	180	10
	3	1	180	4
	4	4	180	1
	5	1	240	10
	6	1	240	10
	7	1	240	45
Flickr	1	16	1500	1
	2	357	240	1
	3	16	60	10
	4	16	480	5
	5	16	20	10
Screen	1	357	30	1
	2	357	80	3
	3	357	180	3
	4	16	20	10
Buscador	1	357	30	1
	2	357	120	1
	3	357	120	3
	4	16	10	10
	5	16	60	10
	6	71	120	5

Tabla 7. Tabla resumen de los distintos experimentos llevados a cabo.

A continuación describiremos las pruebas realizadas para cada uno de los servicios y, posteriormente, pasaremos a visualizar de forma gráfica y analizar los resultados de

dichas pruebas. En las ocasiones en que sea interesante, realizaremos también una comparativa entre servicios para resolver cuestiones sobre si hay más similitudes o distinciones entre ellos en cuanto a la publicidad basada en intereses.

8.1. Experimentos realizados con Correo Yahoo

Este producto Yahoo es, sencillamente, un servicio de correo electrónico. Es por tanto que los experimentos realizados en este servicio consisten en el envío de un correo electrónico de una cuenta Yahoo a otra.

8.1.1. Definición de los experimentos

Con respecto a este servicio Yahoo, debemos tener en cuenta tres aspectos a la hora de definir los experimentos:

1. El cuerpo del mensaje no puede estar vacío, de lo contrario el sistema no coloca etiquetas.
2. Las etiquetas se asignan a la cuenta de correo especificada, no a la sesión del navegador.
3. Entre dos correos hay que dejar transcurrir un tiempo lo suficientemente largo como para que Yahoo no determine que se está enviando *spam* desde la cuenta de usuario.

Teniendo en cuenta estos tres requisitos, hemos definido una batería de experimentos con el servicio Correo Yahoo que consisten en lo siguiente:

- Enviar un email desde una cuenta de Correo Yahoo a otra.
- Para vincular los experimentos con los distintos términos sobre los que hacemos las pruebas, el asunto del correo será, precisamente, uno de los términos escogidos para hacer el experimento.
- El cuerpo del mensaje no puede estar vacío, por lo que será idéntico al asunto.
- La comprobación de la existencia de categorías etiquetadas se realiza un tiempo prudente después de realizar un envío y justo antes de enviar el siguiente correo, si lo hubiese.

Una vez definida la dinámica de los experimentos, hemos ejecutado varios experimentos que han concluido en resultados muy dispares.

8.1.2. Ejecución de los experimentos

Experimento 1

Definición: 5 términos (nivel 6 de la DB), envía un correo cada 3 minutos aproximadamente, 10 repeticiones. Se utiliza una cuenta de correo que no tiene etiquetas asociadas.

Resolución: Con el primer correo aparecen 17 etiquetas que no están temáticamente relacionadas con el término usado.

Experimento 2

Definición: 4 términos (nivel 5 de la DB), envía un correo cada 3 minutos aproximadamente, 10 repeticiones. Se utiliza una cuenta de correo que sí tiene etiquetas asociadas.

Resolución: No aparecen etiquetas nuevas.

Experimento 3

Definición: 1 término, se envía un correo cada 3 minutos aproximadamente, 4 repeticiones. Se utiliza una cuenta de correo que sí tiene etiquetas asociadas.

Resolución: No aparecen etiquetas nuevas.

Experimento 4

Definición: 4 términos (nivel 5 de la DB), 1 repetición, 180s antes de comprobar las etiquetas. Se utiliza una cuenta de correo que no tiene etiquetas asociadas.

Resolución: Con el primer correo aparecen 17 etiquetas que no están temáticamente relacionadas con el término usado. No son las mismas etiquetas que aparecieron en el experimento 1.

Experimento 5

Definición: 1 término, se envía un correo cada 4 minutos aproximadamente, 10 repeticiones. Se utiliza una cuenta de correo que no tiene etiquetas asociadas.

Resolución: No hay etiquetas.

Experimento 6

Definición: 1 término, se envía un correo cada 4 minutos aproximadamente, 10 repeticiones. Se utiliza una cuenta de correo que no tiene etiquetas asociadas.

NOTA: El término utilizado es el mismo que en el experimento 5.

Resolución: Tras el envío del 7º correo aparecen 32 etiquetas. alguna relacionada temáticamente con el contenido; la mayoría, no.

Experimento 7

Definición: 1 término, se envía un correo cada 4 minutos aproximadamente, 45 repeticiones. Se utiliza una cuenta de correo que no tiene etiquetas asociadas.

Resolución: No hay etiquetas.

8.1.3. Observaciones

De los datos obtenidos en esta batería de experimentos, podemos deducir que el sistema de publicidad basado en intereses de Yahoo sí que tiene en cuenta el servicio de correo, dado que en algunos casos sí que aparecen categorías etiquetadas pero, sin embargo, el sistema de etiquetado no funciona muy bien, ya que no puede obtenerse un patrón de comportamiento ni tampoco una lógica en las etiquetas que aparecen para los términos utilizados.

8.2. Experimentos realizados con Flickr

Flickr es una red social de Yahoo que permite almacenar, ordenar, buscar, vender y compartir fotografías o vídeos en línea.

8.2.1. Definición de los experimentos

Para comprobar la influencia de este servicio en la publicidad basada en intereses de Yahoo, hemos diseñado unos experimentos sencillos que consisten en lo siguiente:

- Abrir la página de Flickr.
- Realizar búsquedas en su buscador utilizando como palabras clave los términos seleccionados en cada experimento.

8.2.2. Ejecución de los experimentos

Experimento 1

Definición: 16 términos (nivel 1 de la DB), una búsqueda cada 25 minutos aproximadamente, 1 repetición.

Resolución: No hay etiquetas.

Experimento 2

Definición: 357 términos (todos los términos de la DB), una búsqueda cada 4 minutos aproximadamente, 1 repetición.

Resolución: No hay etiquetas.

Experimento 3

Definición: 16 términos (nivel 1 de la DB), una búsqueda cada 60 segundos aproximadamente, 10 repeticiones.

Resolución: No hay etiquetas.

Experimento 4

Definición: 16 términos (nivel 1 de la DB), una búsqueda cada 8 minutos aproximadamente, 5 repeticiones.

Resolución: No hay etiquetas.

Experimento 5

Definición: 16 términos (nivel 1 de la DB), una búsqueda cada 20 segundos aproximadamente, 10 repeticiones.

Resolución: No hay etiquetas.

8.2.3. Observaciones

De los datos obtenidos en esta batería de experimentos, podemos concluir en que el sistema de publicidad basado en intereses de Yahoo no tiene en cuenta el servicio de Flickr, dado que realizando búsquedas en la propia página con diferentes términos, tiempos entre búsquedas y repeticiones, no hemos conseguido ninguna etiqueta.

8.3. Experimentos realizados con Yahoo Screen

Yahoo Screen es un servicio de alojamiento de vídeos en el que, inicialmente, los usuarios podían subir sus propios vídeos (como en el caso del conocido Youtube), pero que actualmente sólo dispone de vídeos proporcionados por el propio sitio web.

8.3.1. Definición de los experimentos

Como la única interacción que permite hoy por hoy este servicio es la búsqueda de vídeos (aparte de, obviamente, la reproducción), definiremos los experimentos de la siguiente manera:

- Abrir la página de Yahoo Screen.
- Realizar búsquedas en su buscador utilizando como palabras clave los términos seleccionados en cada experimento.

Cabe destacar que no iniciaremos sesión con ninguna cuenta Yahoo antes de realizar las búsquedas para así evitar que las etiquetas se asocien a dicha cuenta, como ocurría en el caso del correo. Las búsquedas serán, por tanto, anónimas.

8.3.2. Ejecución de los experimentos

Experimento 1

Definición: 357 términos (todos los niveles de la DB), una búsqueda cada 30 segundos aproximadamente, 1 repetición.

Resolución: Un total de 313 categorías etiquetadas.

Experimento 2

Definición: 357 términos (todos los niveles de la DB), una búsqueda cada 80 segundos aproximadamente, 3 repeticiones.

Resolución: Un total de 344 categorías etiquetadas.

Experimento 3

Definición: 357 términos (todos los niveles de la DB), una búsqueda cada 3 minutos aproximadamente, 3 repeticiones.

Resolución: Un total de 343 categorías etiquetadas.

Experimento 4

Definición: 16 términos (nivel 1 de la DB), una búsqueda cada 20 segundos aproximadamente, 10 repeticiones.

Resolución: Un total de 6 categorías etiquetadas.

8.3.3. Análisis de los resultados

Como se puede observar de la ejecución de los experimentos, en Yahoo Screen obtenemos un alto índice de categorías etiquetadas, ya sea variando el número de repeticiones o el período de tiempo transcurrido entre búsquedas. Es por esto que haremos un análisis más detallado que en los casos anteriores, apoyándonos en gráficos para evaluar más a fondo los resultados obtenidos.

En un primer caso, vamos a hacer una evaluación de la cantidad de etiquetas que se obtienen al realizar **una única búsqueda, cada 120 segundos, de 357 términos** diferentes.

Para ello nos apoyaremos en un gráfico de barras, representado en la Figura 11, en el que el eje x nos indica la cantidad de categorías que se han etiquetado para cada término, y el eje y nos indica el número de términos en cuestión. A modo de ejemplo, la primera columna representa los 237 términos que no obtuvieron ninguna etiqueta durante el experimento.

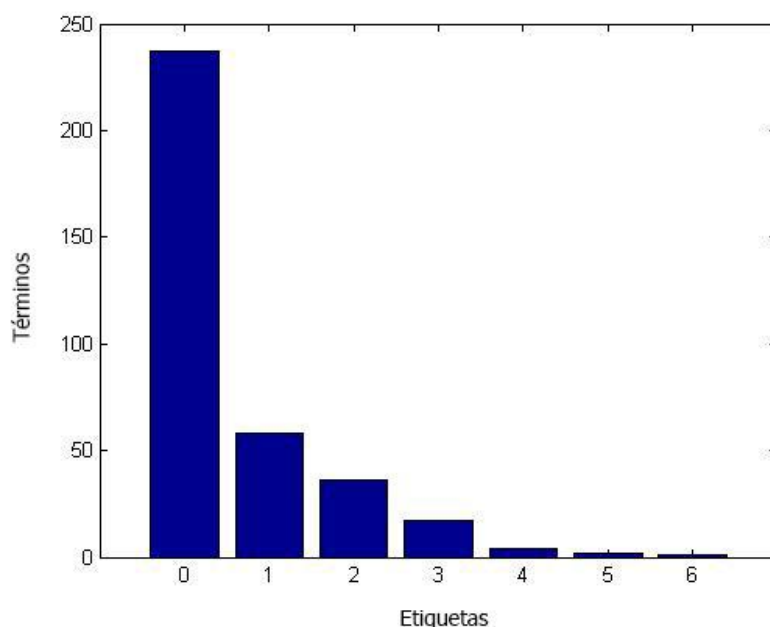


Figura 12. Gráfico de barras que muestra la cantidad de etiquetas que se obtienen al realizar una única búsqueda, cada 120 segundos, de 357 términos diferentes sobre Yahoo Screen.

Como podemos apreciar en la Figura 12, el número de términos que no tienen etiquetas es considerablemente más alto que el resto, que continúa un patrón descendente pero menos diferenciado.

De este análisis obtenemos que es más probable tener pocas -o ninguna- categorías etiquetadas en cualquiera de los términos.

Para confirmar que este comportamiento se mantiene aunque cambiemos de experimento, vemos ahora tres gráficas que representan las tres repeticiones de un mismo experimento: **tres búsquedas de cada término, cada 80 segundos cada una, para un total de 357 términos.**

De la Figura 13 obtenemos la confirmación de que la estructura de las etiquetas, en general, se mantiene. Es cierto que el número de términos con cero etiquetas se reduce y pasa a aumentar el número de etiquetas en las siguientes barras conforme aumentamos la repetición pero, como ya se ha dicho, en general, el comportamiento es el mismo.

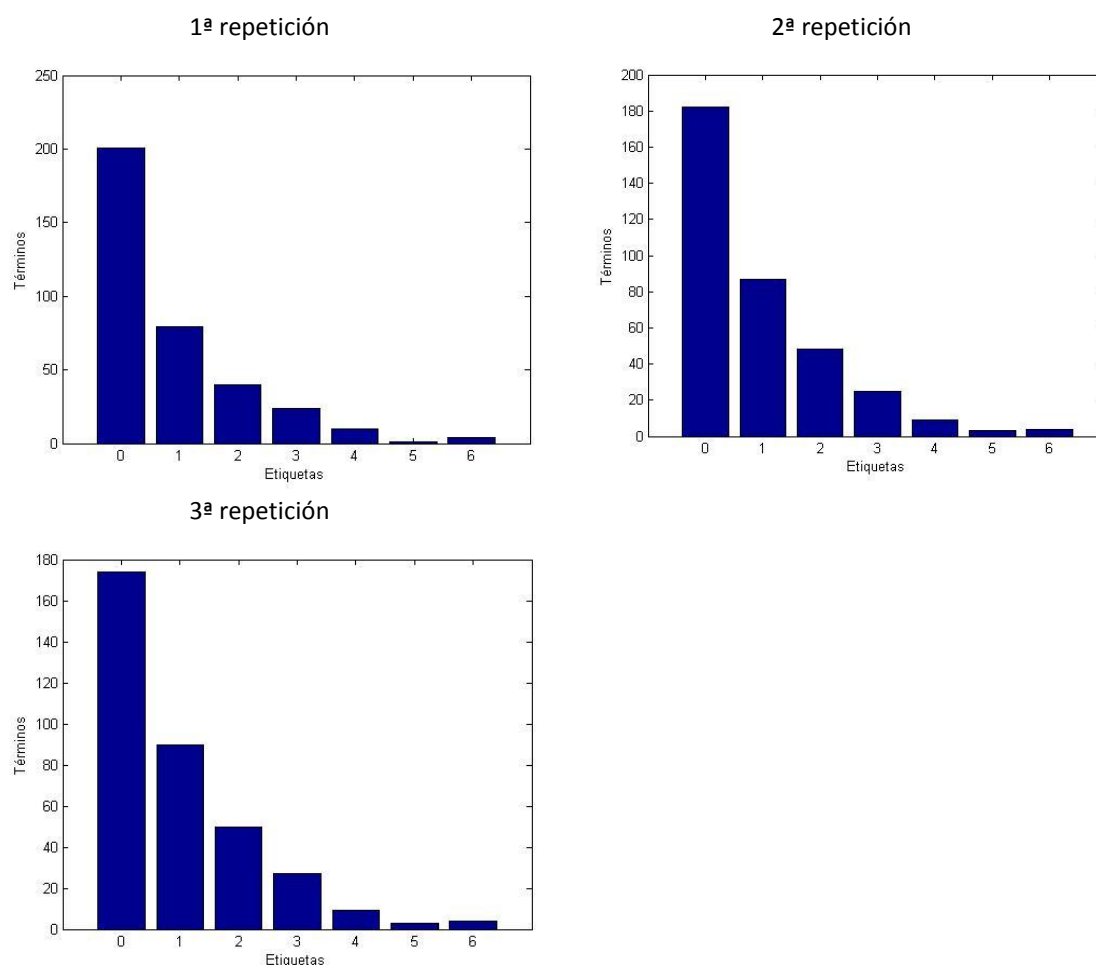


Figura 13. Gráficos de barras que muestran la cantidad de etiquetas que se obtienen al realizar tres repeticiones de un mismo experimento sobre Yahoo Screen: tres búsquedas de cada término, cada 80 segundos cada una, para un total de 357 términos.

Una vez determinado que, aunque variemos el número de repeticiones, la mayor cantidad de términos no consiguen ninguna etiqueta y, tras estos, les siguen por orden los que tienen una etiqueta, después dos, y así sucesivamente, vamos a realizar un estudio sobre el tipo de términos que consiguen -o no- estas etiquetas.

Para esto, hemos seleccionado un conjunto aleatorio de experimentos realizados sobre Yahoo Screen y hemos agrupado los términos en dos bloques:

- Según la temática a la que pertenecen.
- Según el nivel de profundidad en la base de datos.

Realizamos estas distinciones a fin de entender si hay algún tipo de distinción entre términos a la hora de asignar etiquetas o no hacerlo.

El primer análisis lo realizaremos sobre el nivel de categoría de cada etiqueta.

Como ya hemos mencionado anteriormente, la lista de términos con los que realizamos los experimentos, está compuesta sencillamente por cada una del

listado de categorías de Yahoo. Es por esto que las etiquetas corresponden cada una a un nivel distinto en la base de datos en la que las hemos almacenado. Las categorías más generales serán de nivel 1 e iremos incrementando el nivel conforme nos adentremos en categorías más específicas, hasta llegar al nivel 6.

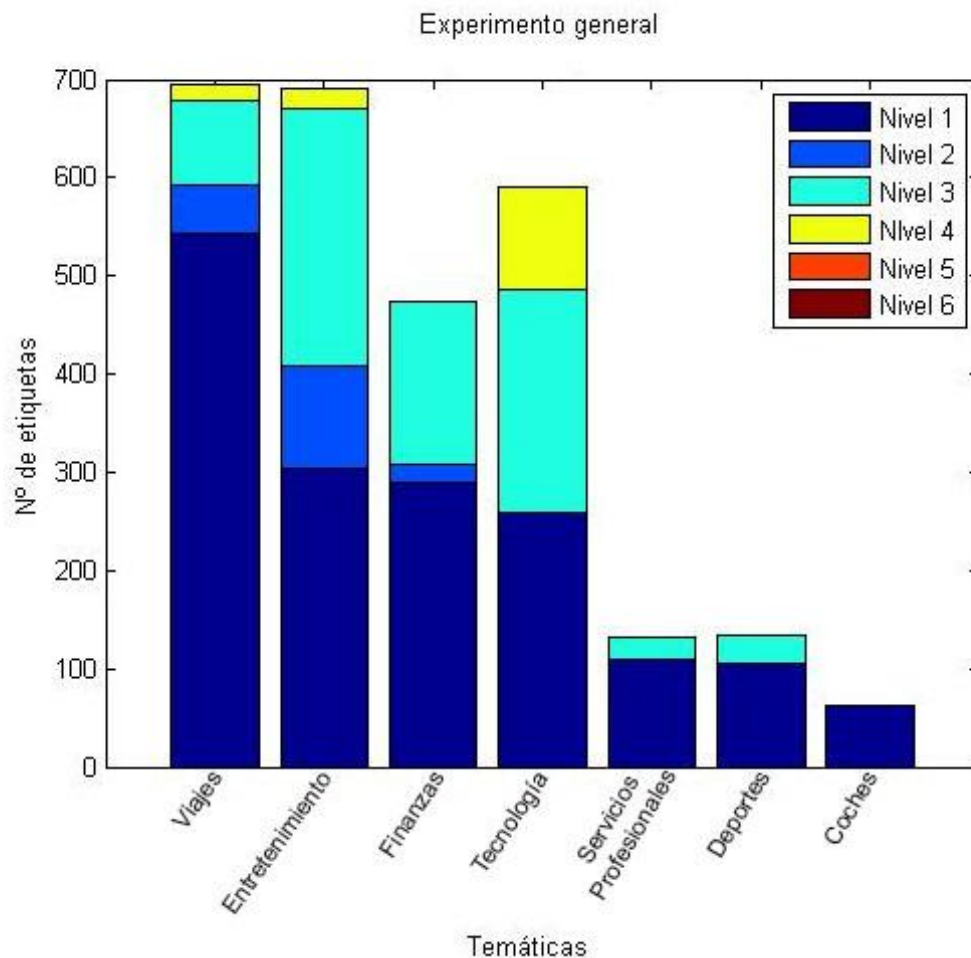


Figura 14. Gráfico de barras por colores que representa la cantidad de etiquetas de cada nivel de categoría que encontramos en un conjunto de experimentos aleatorios sobre Yahoo Screen, con distintos tiempos entre búsquedas y distintas repeticiones, pero siempre utilizando los 357 términos disponibles.

La Figura 14 se ha obtenido a partir de un conjunto de experimentos aleatorios, todos realizados sobre Yahoo Screen, con distintos tiempos entre búsquedas y distintas repeticiones, pero siempre utilizando los 357 términos disponibles.

Para realizar esta representación, hemos agrupado las categorías más etiquetadas en siete grupos que las engloban por temática: Viajes, Entretenimiento, Finanzas, Tecnología, Servicios Profesionales, Deportes y Coches. Es decir, cada una de las etiquetas encontradas pertenecerá a uno de estos grupos.

De esta forma, al igual que en las Figura 12 y Figura 13, se muestra la repartición de etiquetas. No obstante, gracias a la representación por colores, se muestra también a qué grupo pertenece cada una de ellas.

El objetivo de esta representación es obtener una visión global de las etiquetas obtenidas en Yahoo Screen. Por un lado, cada barra representa cada uno de los conjuntos de categorías que más etiquetas consiguen, por lo que podremos ver con facilidad qué tipo son las categorías más etiquetadas. Por otro lado, la división por colores nos permitirá ver el nivel al que pertenecen esas etiquetas, es decir, cómo de generales son (recordemos que el nivel 1 es el más general y, el nivel 6, el más específico).

Atendiendo a todo esto, observamos varias cosas:

- Las etiquetas relacionadas con "Viajes" son las más utilizadas, seguidas muy de cerca por las relacionadas con "Entretenimiento" y, en tercer lugar, las que tiene que ver con "Tecnología".
- El color predominante en la gráfica es el azul marino, luego la cantidad mayoritaria de etiquetas es de nivel 1, por lo que podemos deducir que las categorías más generales se etiquetan con más facilidad que las específicas: los niveles 5 y 6, que son los más específicos, ni siquiera aparecen representados en el gráfico.
- Mientras que todos los niveles cumplen que cuanto más específicos son menos etiquetas aparecen, encontramos una salvedad en las etiquetas de nivel 2 que, a pesar de ser más general que el nivel 3, aparece con mucha menos frecuencia.
- De esta figura podemos deducir también que las etiquetas más específicas que se obtienen son de nivel 4.
- Asimismo, podemos añadir que hay categorías que nunca han aparecido etiquetadas en ninguno de los experimentos realizados.

El último estudio sobre Yahoo Screen lo realizaremos a partir de la Figura 15, que representa los tipos de categorías que más se etiquetan y los tipos de palabras que más etiquetas reciben.

De nuevo, atendiendo a los bloques que comprenden las categorías más etiquetadas y a partir de un experimento representativo, podemos visualizar los dos tipos de informaciones que nos interesan:

- En la barra de la izquierda tenemos un conjunto de las categorías más etiquetadas divididas en grupos por su temática. Podemos ver que es la de "Entretenimiento" la categoría predominante, seguida por las

correspondientes a "Tecnología", "Viajes" y "Finanzas", que tienen cantidades similares de etiquetas.

- En la barra de la derecha hemos representado las palabras a las que se asocian las categorías que, como podemos ver, es más pequeña que la anterior, de lo que se deduce que cada término buscado suele conseguir más de una etiqueta. En este caso es predominante el bloque "Viajes", seguido, esta vez más claramente, por "Finanzas".

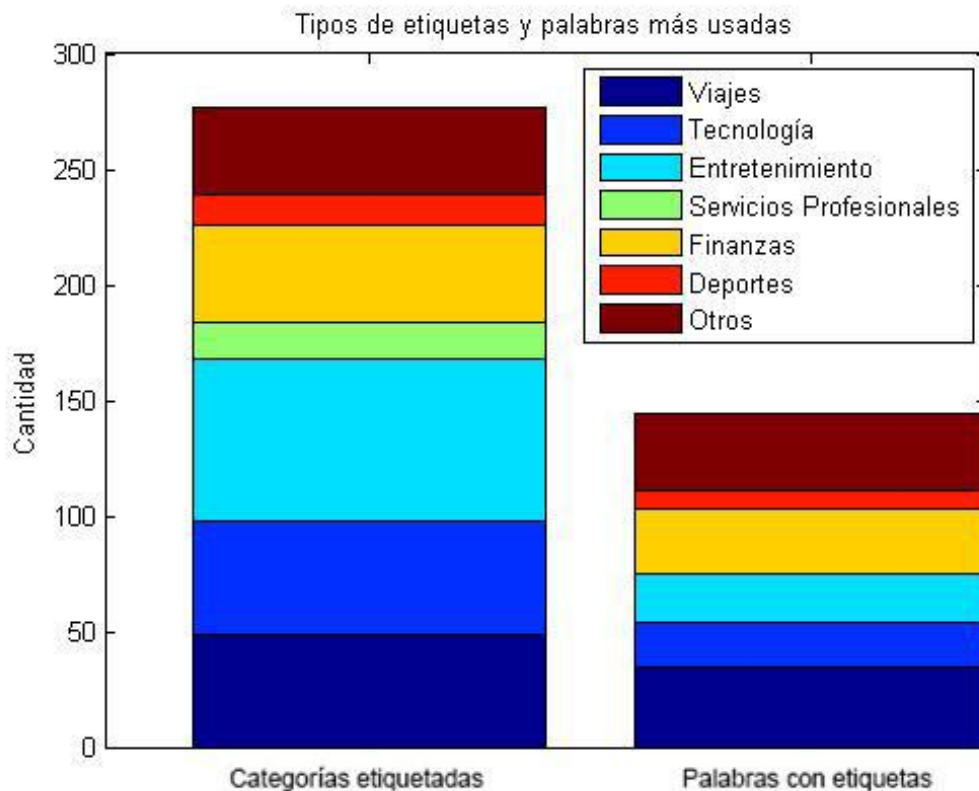


Figura 15. Gráfico de barras por colores que representa la cantidad de etiquetas de cada temática que encontramos en un conjunto de experimentos aleatorios sobre Yahoo Screen. La barra de la izquierda estudia este comportamiento entre las etiquetas conseguidas. La barra de la derecha estudia este comportamiento entre las palabras que tienen etiquetas.

En cuanto a la especificidad de los términos, es decir, lo generales o no que son, tanto de los etiquetados como de los que se utilizan es las búsquedas, podemos determinar ciertos aspectos apoyándonos en la Figura 16.

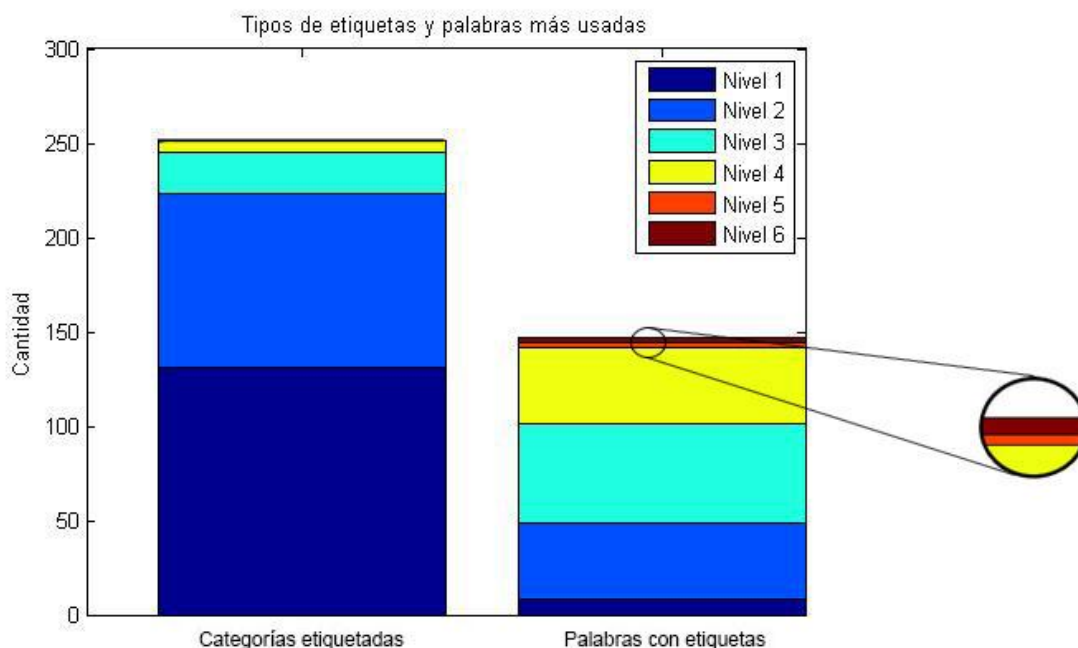


Figura 16. Gráfico de barras por colores que representa la cantidad de etiquetas de cada nivel de categoría que encontramos en un conjunto de experimentos aleatorios sobre Yahoo Screen. La barra de la izquierda estudia este comportamiento entre las etiquetas conseguidas. La barra de la derecha estudia este comportamiento entre las palabras que tienen etiquetas.

Hemos hecho zoom en ambas barras a fin de corroborar que, a pesar de que se trata de una cantidad mínima, a diferencia de la barra para las categorías etiquetadas, sí que encontramos términos muy específicos que consiguen etiquetas, lo cual afirma que un término específico no tiene por qué conseguir una etiqueta tan específica como lo es él.

8.3.4. Observaciones

Del estudio realizado en Yahoo Screen obtenemos varios conceptos claros sobre el sistema de etiquetado de categorías:

1. En general, el número de etiquetas para cada término sigue un esquema constante aunque variemos el número de repeticiones o el tiempo entre búsquedas: hay un mayor número de términos con cero etiquetas, y a este le siguen los que tienen una etiqueta, dos, y así sucesivamente.
2. Es importante la temática del término que buscamos (las palabras relacionadas con "Viajes" son siempre las que más etiquetas consiguen).
3. Se obtienen más fácilmente etiquetas correspondientes a categorías generales (nivel 1) que otras más específicas (nivel 6).
4. Hay categorías que nunca se etiquetan, o bien, es necesario definir otro tipo de experimentos para llegar a ellas.
5. Las palabras que consiguen etiquetas, en general, consiguen más de una.

6. No siempre una etiqueta perteneciente a una determinada categoría general consigue etiquetas pertenecientes a esa categoría, es decir, un término perteneciente a "Entretenimiento" puede conseguir una etiqueta de "Tecnología".

8.4. Experimentos realizados con Buscador Yahoo

Este servicio consiste, como su propio nombre indica, en un servicio de búsquedas en la web realizado por Yahoo.

8.4.1. Definición de los experimentos

En este caso, dado que nos encontramos ante un servicio sobre el que sólo se nos permite hacer búsquedas, los experimentos definidos serán similares a los vistos en Flickr y Yahoo Screen.

- Abrir la página inicial de Yahoo.
- Realizar búsquedas en la barra de buscador, utilizando como palabras clave los términos seleccionados en cada experimento.

En este caso tampoco iniciaremos sesión con ninguna cuenta Yahoo antes de realizar las búsquedas para así evitar que las etiquetas se asocien a dicha cuenta. De nuevo, las búsquedas serán anónimas.

8.4.2. Ejecución de los experimentos

Experimento 1

Definición: 357 términos (todos los niveles de la DB), una búsqueda cada 30 segundos aproximadamente, 1 repetición.

Resolución: Un total de 5 categorías etiquetadas.

Experimento 2

Definición: 357 términos (todos los niveles de la DB), una búsqueda cada 2 minutos aproximadamente, 1 repetición.

Resolución: Un total de 326 categorías etiquetadas.

Experimento 3

Definición: 357 términos (todos los niveles de la DB), una búsqueda cada 2 minutos aproximadamente, 3 repeticiones.

Resolución: Un total de 351 categorías etiquetadas.

Experimento 4

Definición: 16 términos (nivel 1 de la DB), una búsqueda cada 10 segundos aproximadamente, 10 repeticiones.

Resolución: No hay categorías etiquetadas.

Experimento 5

Definición: 16 términos (nivel 1 de la DB), una búsqueda cada 60 segundos aproximadamente, 10 repeticiones.

Resolución: Un total de 2 categorías etiquetadas.

Experimento 6

Definición: 71 términos (nivel 4 de la DB), una búsqueda cada 2 minutos aproximadamente, 5 repeticiones.

Resolución: Un total de 101 categorías etiquetadas.

8.4.3. Análisis de los resultados

Como se puede observar de la ejecución de los experimentos, en el buscador Yahoo obtenemos un alto índice de categorías etiquetadas siempre que el período de tiempo transcurrido entre búsquedas sea suficientemente grande. Digamos que si este tiempo es muy pequeño, el sistema de etiquetas no funciona correctamente.

Como con este servicio también obtenemos un alto número de categorías etiquetadas, realizaremos, como en el caso anterior, un análisis más detallado que en los dos primeros servicios, apoyándonos también en gráficos para evaluar más a fondo los resultados obtenidos.

En la Figura 17, encontramos un gráfico de barras que representa la cantidad de etiquetas obtenidas al realizar **una única búsqueda, cada 120 segundos, de 357 términos** diferentes. El eje x nos indica la cantidad de categorías que se han etiquetado para cada término, y el eje y nos indica el número de términos en cuestión.

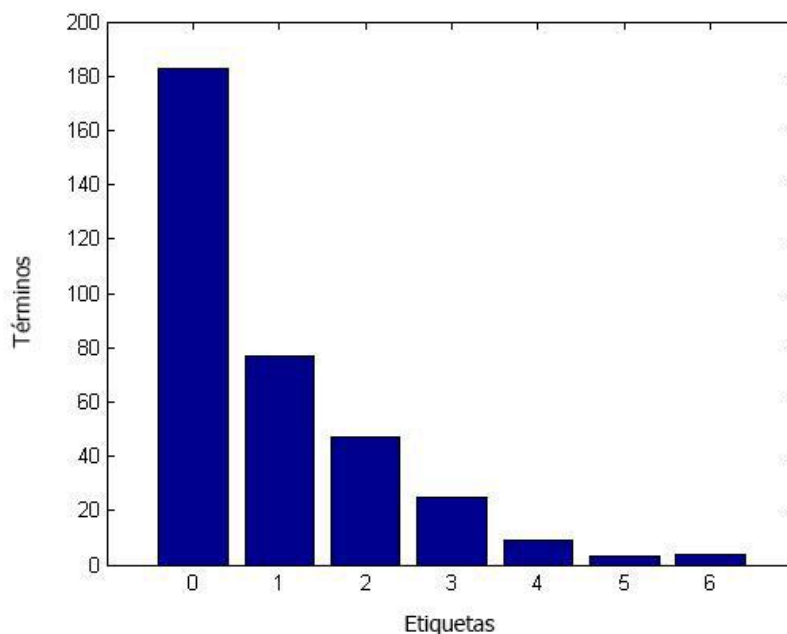


Figura 17. Gráfico de barras que muestra la cantidad de que se obtienen al realizar una única búsqueda, cada 120 segundos, de 357 términos diferentes sobre Buscador Yahoo.

Como podemos observar, la forma del gráfico es muy similar a la obtenida para el mismo experimento en Yahoo Screen (Figura 12): el número de términos que no tienen etiquetas es considerablemente más alto que el resto (aunque ligeramente menor que para Yahoo Screen, ya que se han conseguido menos etiquetas), y éste continúa un patrón descendente pero menos diferenciado.

A simple vista parece que encontraremos similitudes entre estos dos servicios, por lo que más adelante realizaremos una comparativa entre ambos.

Para ver si este patrón se mantiene, variamos el número de repeticiones de las búsquedas de cada término. Observamos el resultado en la Figura 18.

Como es de esperar, el patrón descendente de los gráficos se mantiene en las sucesivas repeticiones del experimento. Cabe mencionar que, de nuevo, el número de términos con cero etiquetas se reduce y pasa a aumentar el número de etiquetas en las siguientes barras conforme aumentamos la repetición.

Visto el comportamiento de la cantidad de etiquetas conseguidas en base a las repeticiones, realizaremos el estudio sobre el tipo de términos que consiguen etiquetas y las temáticas a las que pertenecen estas categorías.

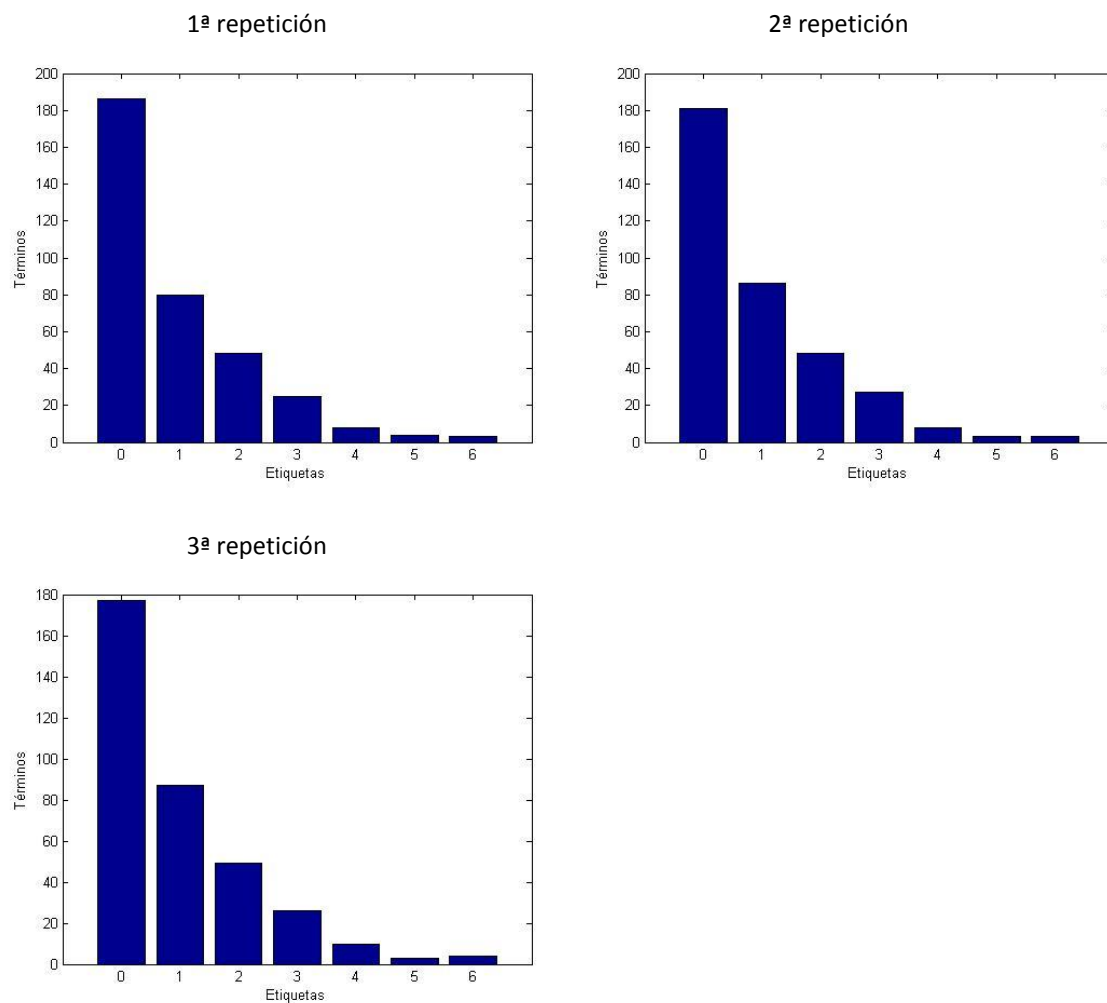


Figura 18. Gráficos de barras que muestran la cantidad de que se obtienen al realizar tres repeticiones de un mismo experimento sobre Buscador Yahoo: tres búsquedas de cada término, cada 120 segundos cada una, para un total de 357 términos.

Como hicimos en el caso anterior, seleccionamos un conjunto aleatorio de experimentos realizados sobre el producto, que en este caso es el buscador de Yahoo, y agrupamos los términos en los dos bloques definidos para Yahoo Screen:

- Según la temática a la que pertenecen.
- Según el nivel de profundidad en la base de datos.

Nos apoyaremos de nuevo en los gráficos de barras de colores para distinguir entre los distintos tipos de categorías y el nivel de especificidad al que pertenece cada una.

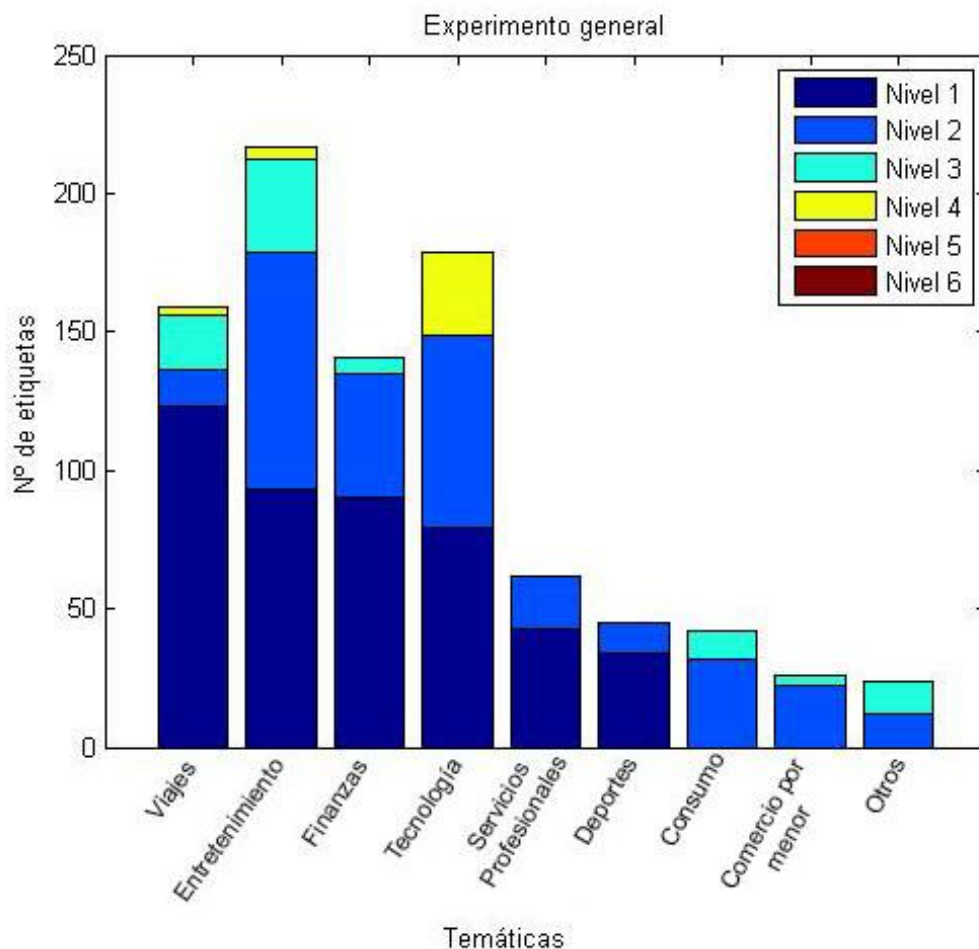


Figura 19. Gráfico de barras por colores que representa la cantidad de etiquetas de cada nivel de categoría que encontramos en un conjunto de experimentos aleatorios sobre Buscador Yahoo, con distintos tiempos entre búsquedas y distintas repeticiones, pero siempre utilizando los 357 términos disponibles.

En el caso de Yahoo Screen, hemos determinado cuáles son los tipos de categorías más etiquetados que, como vemos en la Figura 19, no coinciden exactamente con los que se etiquetan al utilizar el buscador.

Para el buscador, los tipos de categorías que más se etiquetan son, de mayor a menor cantidad, "Entretenimiento", "Tecnología", "Viajes" y "Finanzas". Por detrás de estos, con un número de etiquetas bastante menor, se sitúan "Servicios Profesionales", "Deportes", "Consumo" y "Comercio por menor".

En cuanto a lo específicas que son las categorías etiquetadas, de nuevo comprobamos que la gran mayoría pertenece al nivel 1, que se corresponde con los términos más generales. No obstante, en los grupos de "Entretenimiento" y "Tecnología" se consiguen muchas etiquetas de nivel 2, que, dentro de su generalidad, son algo más específicas que las de nivel 1. Y es precisamente en el grupo "Tecnología" donde encontramos un gran número de etiquetas de nivel 4.

Cabe mencionar que en "Consumo" y "Comercio por menor" no encontramos ninguna etiqueta general, todas corresponden a los niveles 2 y 3.

Así pues, observamos:

- Las etiquetas relacionadas con "Entretenimiento" son las más utilizadas, seguidas por las relacionadas con "Tecnología" y, en tercer lugar, las que tiene que ver con "Viajes".
- El color predominante en la gráfica es el azul marino, luego la cantidad mayoritaria de etiquetas es de nivel 1, por lo que podemos deducir que las categorías más generales se etiquetan con más facilidad que las específicas: los niveles 5 y 6, que son los más específicos, ni siquiera aparecen representados en el gráfico. Con la salvedad de que algunos grupos, como es "Consumo", etiqueta más fácilmente categorías un poco más específicas (nivel2).
- Para la temática "Tecnología" encontramos la salvedad de que existe una gran cantidad de categorías específicas (nivel 4) etiquetadas, en el resto de columnas, mayoritariamente, el número de etiquetas suele descender conforme más específicas son las categorías.
- De esta figura podemos deducir también que las etiquetas más específicas que se obtienen son de nivel 4.
- Asimismo, podemos añadir que hay categorías que nunca han aparecido etiquetadas en ninguno de los experimentos realizados.

A continuación, pasaremos a realizar el estudio sobre los tipos de palabras que más etiquetas reciben, a través de las siguientes figuras.

En la Figura 20 podemos observar cómo de específicos son, tanto las categorías etiquetadas, como los términos que consiguen las etiquetas.

La barra de la izquierda representa la especificidad de las categorías etiquetadas en un conjunto de experimentos representativo de las búsquedas realizadas sobre el producto Buscador Yahoo. La barra de la derecha representa, para el mismo conjunto, la especificidad de los términos que se han buscado y han conseguido al menos una etiqueta.

Como se puede observar en la Figura 20, el número de etiquetas es muchísimo mayor que el de términos con etiquetas, de donde deducimos que, cuando un término consigue etiquetas, serán, en reglas generales, más de una.

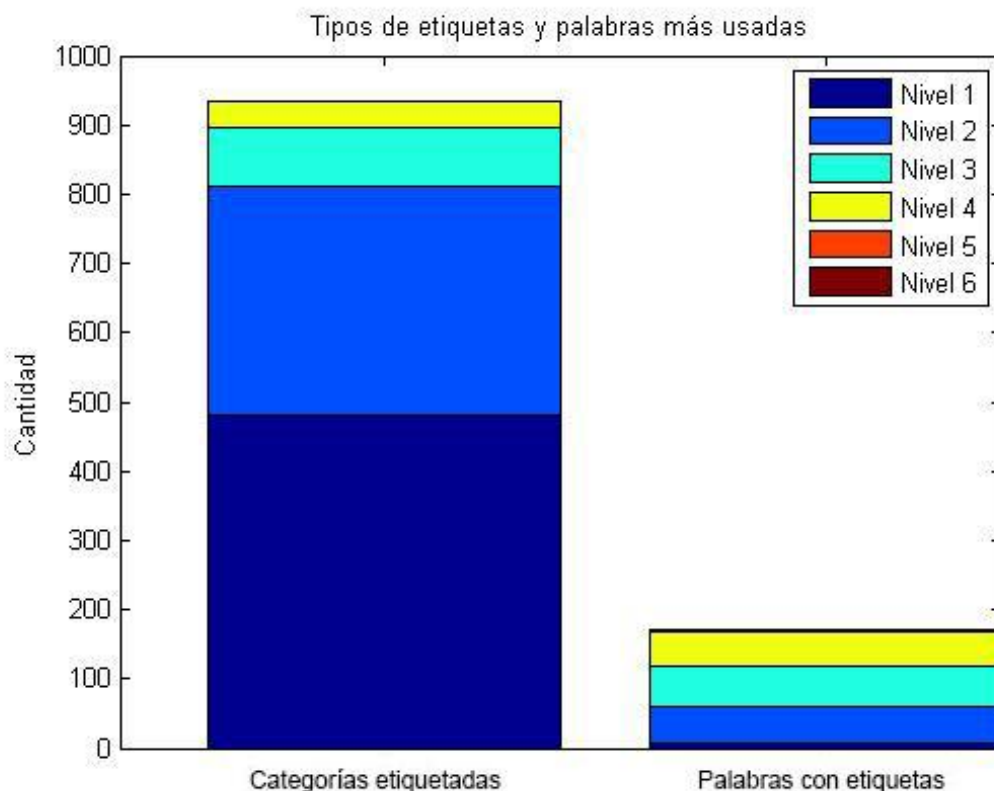


Figura 20. Gráfico de barras por colores que representa la cantidad de etiquetas de cada nivel de categoría que encontramos en un conjunto de experimentos aleatorios sobre Buscador Yahoo. La barra de la izquierda estudia este comportamiento entre las etiquetas conseguidas. La barra de la derecha estudia este comportamiento entre las palabras que tienen etiquetas.

En cuanto a su especificidad, para las categorías etiquetadas vemos que, claramente, predominan las de nivel 1 y nivel dos, es decir, categorías muy generales. Sin embargo, cuando atendemos a los términos buscados a los que se han asociado las etiquetas, los niveles más presentes son el 2, el 3 y el 4. El nivel 1, que es el más general, está casi tan poco presente como los niveles más específicos (5 y 6). En la Figura 21 se representa un zoom de esta columna para apreciar con más precisión los colores.

Por último añadiremos que, mientras que en la columna de la izquierda no encontramos ninguna etiqueta de los niveles más específicos, en la de la derecha sí, lo que quiere decir que términos específicos sí consiguen etiquetas aunque estas sean más generales que los propios términos buscados.

Por último, después de determinar que para los términos buscados se consiguen etiquetas de especificidad intermedia en su mayoría, falta saber de qué tipo son estos términos, es decir, a qué tipo de categoría pertenecen.

La Figura 22 muestra los distintos tipos de categorías etiquetadas para el mismo conjunto de experimentos representativo visto anteriormente, y también el tipo al que pertenecen los términos que reciben las etiquetas.

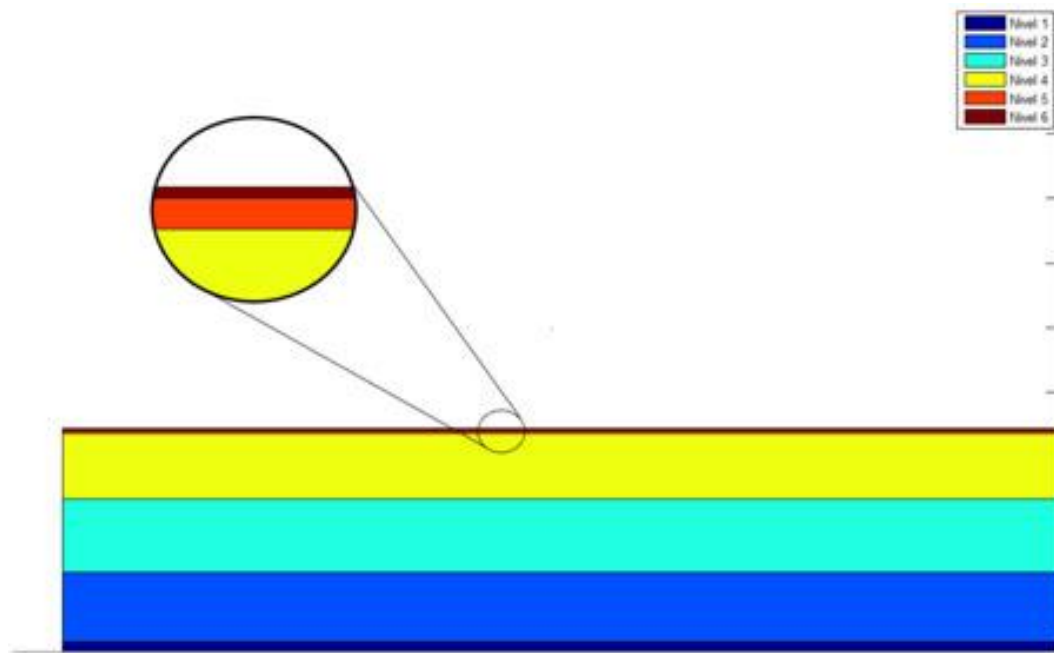


Figura 21. Zoom de la columna perteneciente a las categorías con etiquetas (derecha) de la Figura 20.

La Figura 23 muestra un zoom de la barra correspondiente a las palabras que reciben las etiquetas (derecha) para una mejor visualización de los resultados.

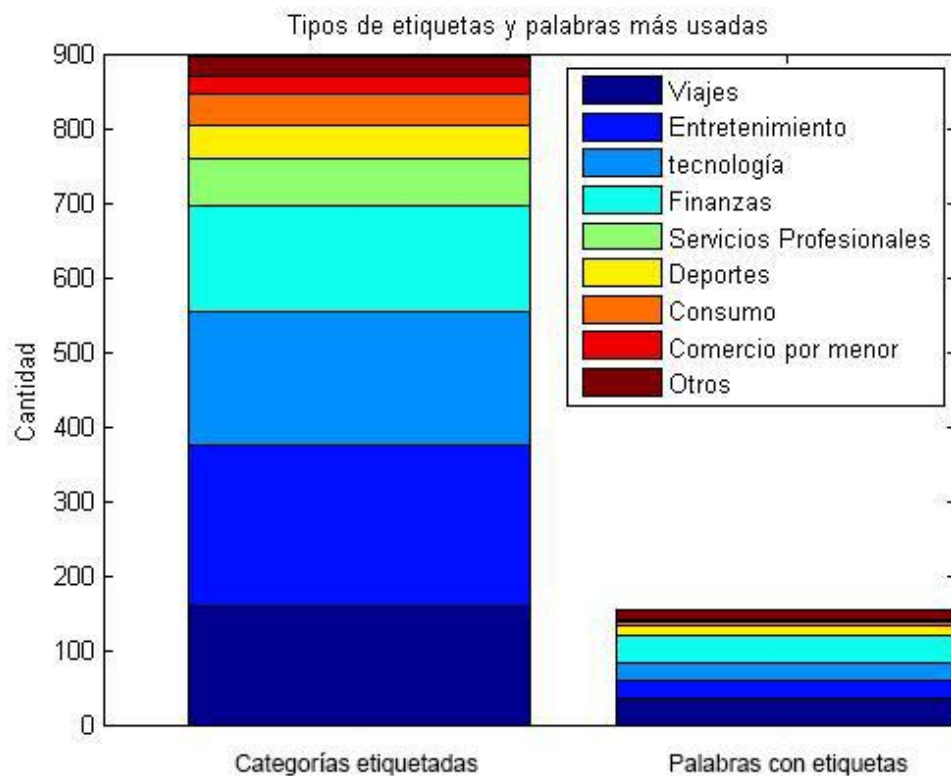


Figura 22. Gráfico de barras por colores que representa la cantidad de etiquetas de cada temática que encontramos en un conjunto de experimentos aleatorios sobre Buscador Yahoo. La barra de la izquierda estudia este comportamiento entre las etiquetas conseguidas. La barra de la derecha estudia este comportamiento entre las palabras que tienen etiquetas.

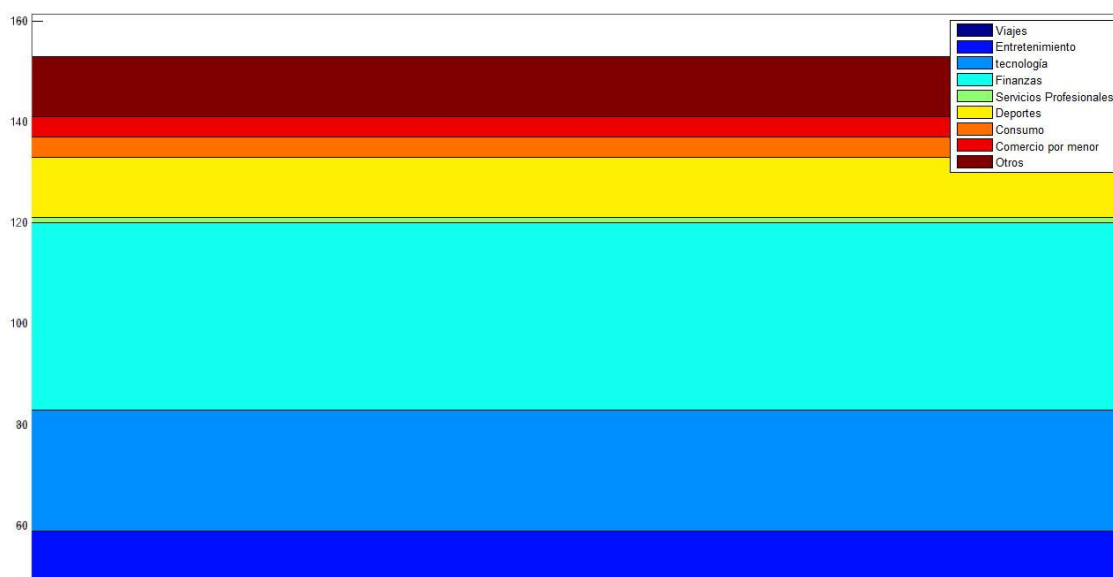


Figura 23. Zoom de la columna perteneciente a las categorías con etiquetas (derecha) de la Figura 22.

De aquí podemos deducir que tipos de categorías como son "Finanzas" o "Viajes", e incluso "Entretenimiento" y "Tecnología" están tan presentes en las categorías etiquetadas como en los términos que consiguen dichas etiquetas. No obstante, se producen diferencias si atendemos al resto de tipos, ya que el hecho de que aparezcan etiquetas relacionadas con un tema en concreto, como puede ser "Servicios Profesionales", no quiere decir que los términos de este grupo consigan etiquetas, como es el caso.

8.4.4. Observaciones

Del estudio realizado en Buscador Yahoo obtenemos varios conceptos claros sobre el sistema de etiquetado de categorías:

1. Si no transcurre el tiempo suficiente entre una búsqueda y la siguiente, el etiquetado no funciona de forma eficiente.
2. En general, el número de etiquetas para cada término sigue un esquema constante aunque variemos el número de repeticiones o el tiempo entre búsquedas: hay un mayor número de términos con cero etiquetas, y a este le siguen los que tienen una etiqueta, dos, y así sucesivamente.
3. Es importante la temática del término que buscamos a la hora de conseguir etiquetas.
4. Se obtienen más fácilmente etiquetas correspondientes a categorías de especificidad intermedia (niveles 3 y 4) frente a otras más generales (nivel1) y las más específicas (niveles 5 y 6).
5. Hay categorías que nunca se etiquetan, o bien, es necesario definir otro tipo de experimentos para llegar a ellas.
6. Las palabras que consiguen etiquetas, en general, consiguen más de una.

7. No siempre una etiqueta perteneciente a una determinada categoría general consigue etiquetas pertenecientes a esa categoría, es decir, un término perteneciente a "Entretenimiento" puede conseguir una etiqueta de "Tecnología".

8.5. Comparativa Buscador y Yahoo Screen

Como hemos podido ver a lo largo de este apartado de pruebas, el funcionamiento del etiquetado entre Yahoo Screen y el buscador, parecen bastante similares.

En realidad, esto es verdad, dado que, si realizamos en ambos productos experimentos similares, podemos constatar que esto es así.

Tomemos como ejemplo un experimento que realiza búsquedas de los 357 términos almacenados en la base de datos, con una única repetición, en la que transcurren unos 100 segundos entre la búsqueda y la comprobación de etiquetas.

Si analizamos las categorías que se etiquetan en uno y en otro, llegamos a la conclusión de que prácticamente todas las categorías que se etiquetan, tanto en un servicio como en el otro, son las mismas, sólo un número mínimo de ellas se diferencian. Si representamos esto en un Diagrama de Venn para visualizar el resultado de forma global, tendríamos algo parecido a lo que muestra la Figura 24.

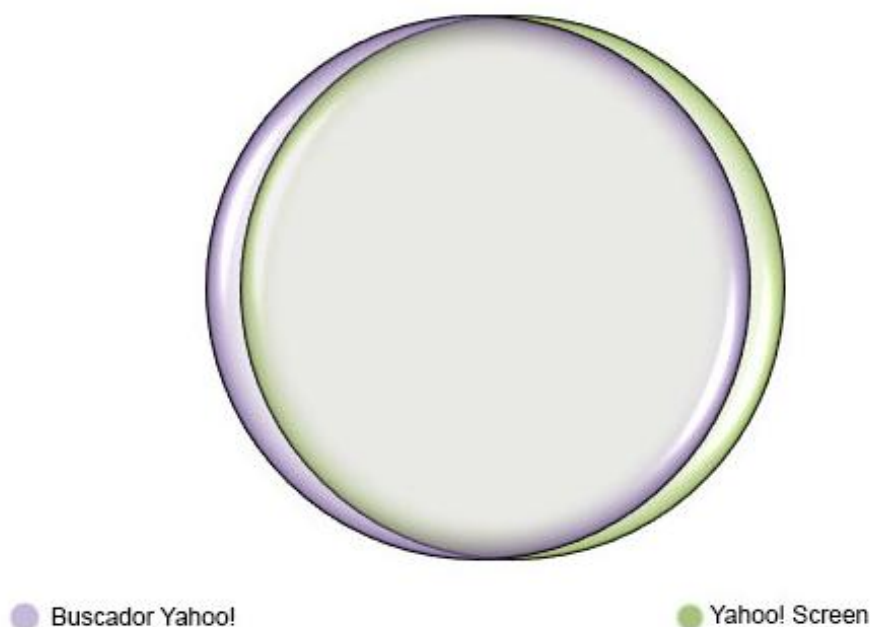


Figura 24. Diagrama de Venn entre las etiquetas obtenidas en Yahoo Screen frente a Buscador Yahoo.

En este Diagrama de Venn, la sección violeta representaría las categorías etiquetadas tras realizar las búsquedas en Buscador Yahoo y, de igual manera, la

sección verde representaría las etiquetadas tras realizar las búsquedas con Yahoo Screen. Prácticamente todas las etiquetas se corresponden en ambos casos, de ahí la zona grisácea, resultante de la confluencia de ambos círculos. Tan sólo unas pocas etiquetas serían independientes para cada producto utilizado.

Si atendemos a qué tipos de categorías aparecen etiquetadas, como ya hemos adelantado en el apartado de Buscador Yahoo, los grupos de "Viajes", "Entretenimiento", "Finanzas" y "Tecnología" se mantienen primeros en el ranking en ambos servicios, sólo que, mientras que en Yahoo Screen teníamos el grupo "Viajes" predominando en la cantidad de etiquetas, en el buscador lo vemos relegado a un tercer puesto, por detrás de "Entretenimiento" y "Tecnología". Otra diferencia notable es que en Buscador Yahoo encontramos, entre las categorías que más aparecen etiquetadas, los grupos "Consumo" y "Comercio por menor", que apenas conseguían etiquetas con Yahoo Screen y que, en este caso, han sustituido al grupo "Coches".

De esta comparativa obtenemos una conclusión más, y es que, al menos para los servicios que funcionan de forma idéntica, existen unas categorías que siempre se etiquetan más que otras, al igual que hay categorías que nunca aparecen etiquetadas.

9. Conclusión

Después de analizar el modelo de publicidad digital que mantenemos en la actualidad, de estudiar las estrategias que se siguen para mantenerlo en desarrollo y sacar el máximo provecho de él, y tras realizar los experimentos necesarios con la herramienta implementada, podemos concluir en varios puntos acerca del modelo de publicidad basada en intereses de Yahoo.

1. En cuanto a si Yahoo se apoya en el uso de sus servicios para determinar los intereses de sus usuarios, definitivamente sí. De los cuatro servicios analizados (Correo, Buscador, Flickr y Screen), hemos obtenido resultados positivos con tres de ellos, así que Yahoo obtiene información de varios de sus productos para elaborar su modelo de publicidad basada en intereses.
2. Siguiendo con el punto anterior, podemos decir que Yahoo no emplea de igual forma los servicios, es decir, el sistema de categorías de intereses no funciona de la misma manera cuando utilizamos uno u otro. Por ejemplo, de los experimentos realizados con el servicio de correo no hemos podido determinar un patrón de comportamiento, ya que el sistema de etiquetas produce resultados impredecibles. Sin embargo, con el buscador sí que podemos predecir algunos resultados.
3. Sobre la predictibilidad de los productos:
 - a. Flickr: no se emplea para el modelo de publicidad basada en intereses.
 - b. Correo: completamente impredecible. No se sigue un patrón o lógica de asignación de etiquetas.
 - c. Buscador: a partir de los resultados obtenidos en los experimentos somos capaces de acercarnos a una predicción de comportamiento del modelo de publicidad basada en intereses.
 - d. Screen: realiza predicciones bastante similares a las del buscador.
4. Atendiendo a los resultados del sistema de predicción de intereses:
 - a. De los únicos servicios de Yahoo que conseguimos datos realmente útiles son el buscador y Yahoo Screen.
 - b. De estos podemos decir que funcionan de forma similar aunque las etiquetas que se consiguen, es decir, la predicción de intereses que se le hace a los usuarios, no es exactamente igual en ambos casos. De hecho, cuando realizamos el mismo experimento en cada uno de estos servicios, aunque prácticamente todas las categorías etiquetadas son las mismas y se producen en casi los mismos términos de búsqueda, siempre hay alguna que difiere.
 - c. En general, el número de etiquetas para cada término sigue un esquema constante aunque variemos el número de repeticiones o el tiempo entre búsquedas.

- d. Es importante la temática del término que buscamos a la hora de conseguir etiquetas, es decir, en función del tema al que pertenezca el término buscado, obtendremos más o menos etiquetas.
- e. Cuando se le asignan categorías de intereses a un usuario, normalmente no son muy generales, es decir, el sistema sabe centrarse un poco en los gustos, pero tampoco demasiado específicas.
- f. Hay categorías que nunca se etiquetan, o bien, es necesario definir otro tipo de experimentos para llegar a ellas.
- g. Si un término buscado consigue alguna etiqueta, en la mayoría de los casos serán más de una, especialmente en aquellos temas que más fácilmente consiguen ser etiquetados.
- h. El hecho de buscar un término de un tema determinado no implica que vaya a conseguir etiquetas relacionadas. Puede conseguir etiquetas, por supuesto, pero no necesariamente del tema que debieran (aunque esto se produce en una minoría de casos).
- i. Particularizando para el caso de Buscador Yahoo, si no transcurre el tiempo suficiente entre una búsqueda y la siguiente, el etiquetado no funciona de forma eficiente.

En general, a través de este estudio y gracias a los experimentos realizados, podemos concluir en que el modelo de publicidad basado en intereses de Yahoo aún está en una fase de maduración, es decir, el sistema funciona y, en general, funciona bastante bien: por ejemplo, si hacemos búsquedas sobre países es capaz de detectar que nos interesan los viajes, e incluso se aventura a llegar hasta el continente o el país en las etiquetas que nos asigna.

No obstante, el hecho de que haya servicios para los que no funciona o servicios para los que no muestra resultados coherentes, son objeto de mejora del modelo o, tal vez, simplemente aún están en fase de desarrollo. Sería interesante poder realizar estos mismos experimentos para el resto servicios de Yahoo que, lamentablemente, no han tenido cabida en este proyecto.

A parte de esto, en cuanto al sistema de predicción en sí, podría afinarse más esta predicción y evitar que se etiqueten categorías de interés que en realidad no le interesan al usuario, es decir, para un mejor funcionamiento, podrían evitarse los conflictos entre términos, aunque sabemos que es algo realmente difícil, pues en muchas ocasiones existen conceptos muy subjetivos y una búsqueda verdaderamente no tiene por qué reflejar nuestro verdadero interés.

Así que, para finalizar este estudio, diremos que Yahoo tiene un sistema bastante potente de predicción de intereses del que estamos seguros que, con un poco más de desarrollo, sería capaz de mostrar a sus usuarios anuncios publicitarios de temas que sean, en su totalidad, interesantes para cada uno de ellos.

ANEXO I

En este anexo se listan las tareas llevadas a cabo por el bloque de control y procesado, que gestiona el funcionamiento general de la herramienta mediante Shell Scripting:

- Control de los archivos de salida y error de la aplicación.
- Conexión y acceso a la base de datos.
- Gestión de la carpeta que guarda el *backup* de los datos.
- Gestión de los términos de búsqueda.
- Control de carpetas y archivos necesarios para realizar los distintos procesos.
- Procesado de archivos.
 - Eliminación de caracteres inservibles y líneas repetidas.
 - Reordenación de datos en los documentos.
 - Conversión de formatos.
- Manejo y control del acceso y la salida del bloque de automatización.
- Gestión de tiempos.
- Liberación de recursos.
- Generación y preparación de documentos previa al almacenaje en la base de datos.
- Control de posibles *bugs* en los experimentos.
- Creación del archivo *zip* que contiene los datos de *backup*.

ANEXO II

En este anexo se listan las tareas llevadas a cabo por el bloque de automatización, en concreto de las tareas programadas en código Javascript, que es el encargado de gestionar las iMacros que automatizan las actividades en el navegador.

- Lanzamiento de iMacros.
 - Emisión de mensajes sobre la actividad del navegador.
 - Borrado de memoria caché.
 - Lectura de las opciones de configuración.
 - Control de fecha y hora.
 - Automatización de las búsquedas.
 - Comprobación de etiquetas.
 - Extracción de información relevante en documentos de texto.
- Control de tiempos.
- Control de repeticiones.
- Control del término de búsqueda empleado en cada momento.

ANEXO III

En este anexo se encuentra disponible el código perteneciente a las iMacros utilizadas en el bloque de automatización para realizar las búsquedas en los servicios de Buscador Yahoo y de Correo Yahoo, mostrados a fin de examinar las diferencias entre los servicios en los que meramente se realiza una búsqueda y el que necesita enviar correos electrónicos.

También se muestra como ejemplo una función iMacros que extrae contenido del Ad Interest Manager y otra de cómo se realiza el guardado de dichos datos.

searchingBuscador.iim

VERSION BUILD=7500718 RECORDER=FX

TAB T=1

TAB CLOSEALLOTHERS

URL GOTO=https://es.yahoo.com

SET !VAR1 "Buscador"

'PROMPT {{!VAR1}}

ADD !EXTRACT {{!VAR1}}

SET !VAR2 {{st}}

TAG POS=1 TYPE=INPUT:TEXT ATTR=ID:UHSearchBox CONTENT={{!VAR2}}

TAG POS=1 TYPE=INPUT:SUBMIT FORM=ID:UHSearch ATTR=ID:UHSearchWeb

'Extraer término buscado

ADD !EXTRACT {{!VAR2}}

'PROMPT {{!EXTRACT}}

searchingCorreo.iim

VERSION BUILD=8890130 RECORDER=FX

TAB T=1

URL GOTO=https://login.yahoo.com/config/mail?&.src=ym&.intl=es

TAG POS=1 TYPE=INPUT:TEXT FORM=ID:mbr-login-form ATTR=ID:login-username

CONTENT=proyectoads1@yahoo.com

SET !ENCRYPTION NO

TAG POS=1 TYPE=INPUT:PASSWORD FORM=ID:mbr-login-form ATTR=ID:login-passwd

CONTENT=YahooAds2014

TAG POS=1 TYPE=BUTTON FORM=ID:mbr-login-form ATTR=ID:login-signin

SET !VAR1 "Correo"

ADD !EXTRACT {{!VAR1}}

```
SET !VAR2 {{st}}
TAG POS=1 TYPE=BUTTON ATTR=TITLE:Redactar
WAIT SECONDS=10
TAG POS=1 TYPE=INPUT:TEXT ATTR=ID:to-field CONTENT=proyectoads@yahoo.com
TAG POS=1 TYPE=INPUT:TEXT ATTR=ID:subject-field CONTENT={{!VAR2}}
WAIT SECONDS=10
TAG POS=1 TYPE=A ATTR=TITLE:"Enviar este correo [Ctrl+Intro]"
WAIT SECONDS=10
TAG POS=1 TYPE=A ATTR=ID:yucs-signout
```

extractVisibles.iim

```
VERSION BUILD=10022823
TAB T=1
TAB CLOSEALLOTHERS
URL GOTO=https://info.yahoo.com/privacy/es/yahoo/opt_out/targeting/details.html
```

'Extrae las categorías visibles

```
TAG POS={{alooop}} TYPE=TD ATTR=* EXTRACT=TEXT
```

saveData.txt

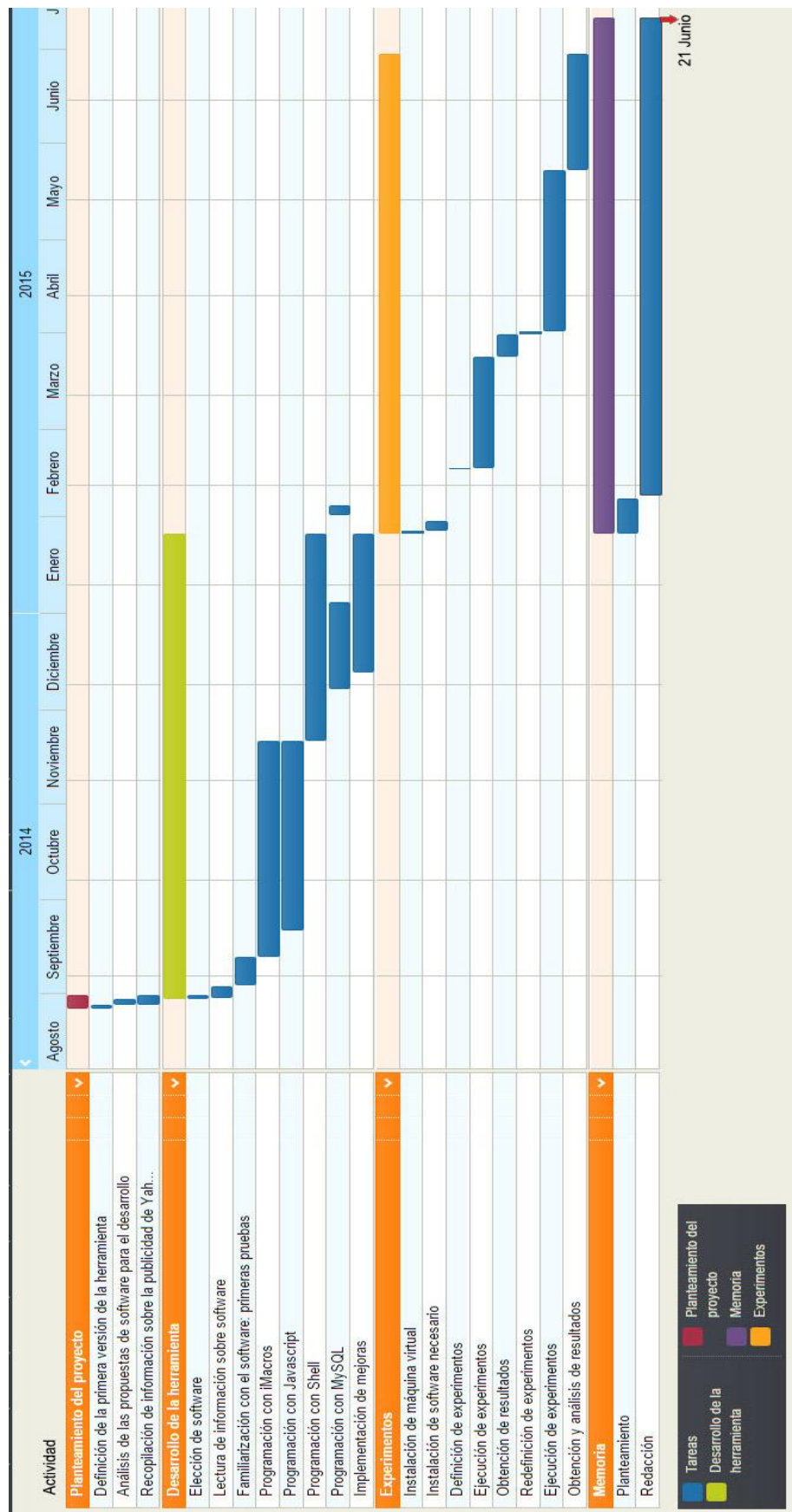
```
VERSION BUILD=10022823
SET !VAR1 {{myDataG}}
SET !EXTRACT {{!VAR1}}
'PROMPT{{!VAR1}}
SAVEAS TYPE=EXTRACT FOLDER=* FILE={{myName}}.txt
```

ANEXO IV

Ejemplo de archivo de configuración.

```
10;      //Repeticiones
1800;    //Duración entre búsquedas en segundos
correo;  //Producto Yahoo sobre el que se realiza la búsqueda (en minúsculas)
0;       //Nivel de categoría (0 - 6). Con 0 seleccionamos todas.
```

ANEXO V



ANEXO VI

En este anexo se detalla el presupuesto que conllevaría el desarrollo de este proyecto en un marco laboral en la actualidad.

En él consta:

- Contratación de un ingeniero Junior para la realización del proyecto.
- Contratación de un ingeniero Sénior para el asesoramiento y supervisión del proyecto.
- Equipos necesarios para la realización del proyecto.
- Licencia de software MATLAB.

Personal	Salario mensual (bruto)	Nº de empleados	Meses contratados	Importe
Ingeniero Junior	1.430,00 €	1	10	14.300,00 €
Ingeniero Sénior	2.140,00 €	1	10	21.400,00 €

Material/Equipo	Coste por unidad	Nº de unidades	Meses alquiler	Importe
Alquiler servidor	72,60 €	1	7	508,20 €
PC de sobremesa	1.000,00 €	1	--	1.000,00 €
Licencia MATLAB	69,00 €	1	--	69,00 €

TOTAL
37.277,20 €

NOTA: El resto de los programas utilizados para el desarrollo del proyecto son de software libre, por lo que no ha sido necesaria la compra de más licencias.

Referencias

- [1] Wikipedia, «Wikipedia,» [En línea]. Available: http://es.wikipedia.org/wiki/Sociedad_de_la_informaci%C3%B3n.
- [2] «Wikipedia,» [En línea]. Available: http://es.wikipedia.org/wiki/Publicidad#La_publicidad_y_los_estereotipos_sociales.
- [3] Wishpond, «Marketing Directo,» [En línea]. Available: <http://www.marketingdirecto.com/actualidad/infografias/18-estadisticas-sobre-publicidad-online-con-las-que-comprendera-que-la-era-2-0-ha-llegado-para-quedarse/>.
- [4] T. d. Miguel y T. de Miguel, «Fundación General CSIC,» [En línea]. Available: http://www.fgcsic.es/lychnos/es_es/articulos/la_internet_del_futuro_y_la_id.
- [5] Wikipedia. [En línea]. Available: http://es.wikipedia.org/wiki/Sociedad_de_consumo.
- [6] V2p, «V2p Online,» [En línea]. Available: <http://www.v2p-online.es/2011/10/15/definicion-publicidad-online/>.
- [7] J. F. Castaño, «platafor.ma,» [En línea]. Available: <http://www.platafor.ma/que-hacemos/publicidad-digital/>.
- [8] J. M. Jiménez, «3Cero,» [En línea]. Available: <http://3cero.com/conceptos-basicos-publicidad-online/>.
- [9] Wikipedia, «Wikipedia,» [En línea]. Available: <https://es.wikipedia.org/wiki/Adserver>.
- [10] Wikipedia, «Wikipedia,» [En línea]. Available: https://en.wikipedia.org/wiki/Ad_exchange.
- [11] «Wikipedia,» [En línea]. Available: https://en.wikipedia.org/wiki/Advertising_network.
- [12] «Basimos,» [En línea]. Available: <http://www.basimos.com/web-marketing/internet-marketing-glossary-c.htm>.
- [13] F. GV, «Wordpress,» Marzo 2014. [En línea]. Available: <https://marketingbuilding.wordpress.com/2014/03/12/online-advertising-publicidad-por-internet/>.
- [14] «Google,» [En línea]. Available: <https://support.google.com/adsense/answer/113771?hl=es>.
- [15] Avocarrot, «Idea To Appster,» [En línea]. Available: <http://www.ideatoappster.com/wp-content/uploads/2013/11/InterestBasedMobileAdvertising.png>.

- [16] L. OPOCE, «Universitat Politècnica de Catalunya,» 2000. [En línea]. Available: http://www.cs.upc.edu/~nicos/proteccion_datos_europa.pdf.
- [17] E. Bru Cuadrada, «Universitat Oberta de Catalunya,» Septiembre 2007. [En línea]. Available: <http://www.uoc.edu/idp/5/dt/esp/bru.pdf>.
- [18] Facebook, «Facebook,» Noviembre 2014. [En línea]. Available: <https://www.facebook.com/business/news/Actualizaciones-a-nivel-mundial-mejoras-en-los-anuncios-de-Facebook-y-presentacin-de-nuevos-controles>.
- [19] G. AdSense, «Google AdSense,» [En línea]. Available: <http://www.google.com/adsense/start/>.
- [20] M. Garcimartin, «Media-tics,» Septiembre 2014. [En línea]. Available: <http://www.media-tics.com/noticia/3787/internet/google-revolucionara-la-publicidad-dirigida-con-un-sistema-propio.html>.
- [21] Yahoo, «Preguntas frecuentes de Yahoo Ad Interest Manager,» [En línea]. Available: <https://policies.yahoo.com/ie/es/yahoo/privacy/topics/opt-outfaq/index.htm>.
- [22] Yahoo, «Política de privacidad de Yahoo,» Julio 2013. [En línea]. Available: <https://policies.yahoo.com/co/es/yahoo/privacy/index.htm>.
- [23] Yahoo, «Ad Interest Manager,» [En línea]. Available: https://info.yahoo.com/privacy/es/yahoo/opt_out/targeting/details.html.
- [24] «NAI Page,» [En línea]. Available: <http://www.networkadvertising.org/choices/>.
- [25] «Yahoo Location,» [En línea]. Available: <https://policies.yahoo.com/us/es/yahoo/privacy/topics/location/index.htm>.
- [26] «iMacros.net,» [En línea]. Available: <http://imacros.net/overview>.
- [27] G. Conde, «Enunmomentillo,» [En línea]. Available: <http://www.enunmomentillo.es/imacros-automatizando-tareas-y-pruebas-en-entornos-web/>.
- [28] «Wikipedia,» [En línea]. Available: https://es.wikipedia.org/wiki/Int%C3%A9rprete_de_comandos.
- [29] «Wikipedia,» [En línea]. Available: http://es.wikipedia.org/wiki/Shell_de_Unix.
- [30] «Wikipedia,» [En línea]. Available: https://es.wikipedia.org/wiki/Base_de_datos.
- [31] «Wikipedia,» [En línea]. Available: <http://es.wikipedia.org/wiki/MySQL>.
- [32] «Yahoo,» [En línea]. Available:

http://info.yahoo.com/privacy/es/yahoo/opt_out/targeting/asc/details.html.

[33] L. O. d. P. O. d. I. C. Europeas, «Universitat Politècnica de Catalunya,» 2000. [En línea]. Available: http://www.cs.upc.edu/~nicos/proteccion_datos_europa.pdf.

[34] Wikipedia, «Wikipedia,» [En línea]. Available: http://es.wikipedia.org/wiki/Publicidad#La_publicidad_y_los_estereotipos_sociales.

[35] «Wikipedia,» [En línea]. Available: <http://es.wikipedia.org/wiki/MySQL>.